Supplementary Information

Inconsistency in large pharmacogenomic studies

Benjamin Haibe-Kains, Nehme El-Hachem, Nicolai Juul Birkbak, Andrew C. Jin, Andrew H. Beck, Hugo J.W.L. Aerts, John Quackenbush

Contents

1 List of Abbreviations

AUC	Area under the drug sensitivity curve.
CGP	Cancer Genome Project initiated by the Wellcome Sanger Institute.
CCLE	The Cancer Cell Liners Encyclopedia initiated by Novartis and the Broad Institute.
IC_{50}	Concentration in micro molar [μM] at which the drug inhibited 50% of the cellular growth.
FDR	False Discovery Rate
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis.
GSK	Cancer cell line screening initiated by GlaxoSmithKline.
R_s	Spearman correlation coefficient

2 Full Reproducibility of the Analysis Results

We will describe how to fully reproduce the figures and tables reported in the main manuscript. We automated the analysis pipeline so that minimal manual interaction is required to reproduce our results. To do this, one must simply:

1. Set up the software environment
2. Run the R scripts
3. Generate the Supplementary Information

2.1 Set up the software environment

We developed and tested our analysis pipeline using R running on linux and Mac OSX platforms.

To mimic our software environment the following R packages should be installed:

- R version 3.0.1 (2013-05-16), `x86_64-unknown-linux-gnu`
- Base packages: base, datasets, graphics, grDevices, grid, methods, parallel, splines, stats, utils
- Other packages: amap 0.8-7, Biobase 2.20.0, BiocGenerics 0.6.0, colorspace 1.2-2, GSA 1.03, MASS 7.3-26, plotrix 3.4-7, prodlim 1.3.7, survcomp 1.10.0, survival 2.37-4, vcd 1.2-13, WriteXLS 2.3.1, xtable 1.7-1
- Loaded via a namespace (and not attached): bootstrap 2012.04-0, epibasix 1.3, KernSmooth 2.23-10, rmeta 2.16, SuppDists 1.1-9, survivalROC 1.0.3, tools 3.0.1

All these packages are available on CRAN[1] or Bioconductor[2], except for jetset which is available on the CBS website[3].

Run the following commands in a R session to install all the required packages:

```
source("http://bioconductor.org/biocLite.R")
biocLite(c("AnnotationDbi", "affy", "affyio", "hthgu133acdf",
 "hthgu133afrmavecs", "hgu133plus2cdf", "hgu133plus2frmavecs",
 "org.Hs.eg.db", "genefu", "biomaRt", "frma", "Hmisc", "vcd",
 "epibasix", "amap", "gdata", "WriteXLS", "xtable", "plotrix",
 "R.utils", "DBI", "GSA", "gplots"))
```

Note that you may need to install Perl[4] and its module Text::CSV_XS for the WriteXLS package to write xls file; once Perl is installed in your system, use the following command to install the Text::CSV_XS module through CPAN[5]:

```
cpan Text/CSV_XS.pm
```

Lastly, follow the instructions on the CBS website to properly install the jetset package or use the following commands in R:

```
download.file(url="http://www.cbs.dtu.dk/biotools/jetset/current/jetset_1.4.0.tar.gz",
 destfile="jetset_1.4.0.tar.gz")
install.packages("jetset_1.4.0.tar.gz", repos=NULL, type="source")
```

[1] http://cran.r-project.org
[2] http://www.bioconductor.org
[3] http://www.cbs.dtu.dk/biotools/jetset/
[4] http://www.perl.org/get.html
[5] http://www.cpan.org/modules/INSTALL.html

Once the packages are installed, uncompress the archive provided as **Supplementary data** accompanying the manuscript[6]. This should create a directory on the file system containing the following files:

`CDRUG_foo.R` Script containing the definitions of all functions required for the analysis pipeline.

`CDRUG_normalization_cgp.R` Script to curate, annotate and normalize of CGP data.

`CDRUG_normalization_ccle.R` Script to curate, annotate and normalize of CCLE data.

`CDRUG_normalization_gsk.R` Script to curate, annotate and normalize of GSK data.

`CDRUG_format.R` Script to identify common cell lines, tissue types and drugs investigated both in CGP and CCLE.

`CDRUG_analysis.R` Script generating all the figures and tables reported in the manuscript.

`CDRUG_analysisbis.R` Script generating Figure 4 in the manuscript.

`CDRUG_analysisbis_gsk.R` Script comparing the IC_{50} measures between CGP, CCLE and GSK.

`CDRUG_pipeine.R` Master script running all the scripts listed above to generate the analysis results.

`gsea2-2.0.13.jar` GSEA java executable; it can also be downloaded from the GSEA website[7].

`c5.all.v4.0.entrez.gmt` Definition of genesets based on Entrez Gene IDs; it can also be downloaded from the GSEA website[8].

`matching_cell_line_CCLE_CGP.csv` Curation of cell line name to match CGP and CCLE nomenclatures.

`matching_tissue_type_CCLE_CGP.csv` Curation of tissue type name to match CGP and CCLE nomenclatures.

`matching_cell_line_GSK_CCLE_CGP.csv` Curation of cell line name to match those of GSK with those of CGP and CCLE.

`matching_tissue_type_GSK_CCLE_CGP.csv` Curation of tissue type name to match those of GSK with those of CGP and CCLE.

`cdrug_suppl_info.tex` The LaTeX file of the present supplementary information

All the files required to run the automated analysis pipeline are now in place. It is worth noting that raw gene expression and drug sensitivity data are voluminous, please ensure that at least 25GB of storage are available.

2.2 Run the R scripts

Open a terminal window and go to the `CDRUG` directory. You can easily run the analysis pipeline either in batch mode or in a R session. Before running the pipeline you can specify the number of CPU cores you want to allocate to the analysis (by default only 1 CPU core will be used). To do so, open the script `CDRUG_pipeline.R` and update line #33:

```
nbcore <- 4
```

[6] The code is also available on GitHub within the **cdrug** repository.
[7] http://www.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/software/gsea2-2.0.13.jar
[8] http://www.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/msigdb/4.0/c5.all.v4.0.entrez.gmt

to allocate four CPU cores for instance.

To run the full pipeline in batch mode, simply type the following command:

```
R CMD BATCH CDRUG_pipeline.R Rout &
```

The progress of the pipeline could be monitored using the following command:

```
tail -f Rout
```

To run the full analysis pipeline in an R session, simply type the following command:

```
source("CDRUG_pipeline.R")
```

Key messages will be displayed to monitor the progress of the analysis.

The analysis pipeline was developed so that all intermediate analysis results are saved in the directories `data` and `saveres`. Therefore, in case of interruption, the pipeline will restart where it stopped.

2.3 Generate the Supplementary Information

After completion of the analysis pipeline a directory `saveres` will be created to contain all the intermediate results, tables and figures reported in the main manuscript and this Supplementary Information.

3 Comparison of experimental protocols

A major potential source of variability in phenotypic measurements between CGP and CCLE phenotype is due to differences in procedures used for growing cells, storing compounds, treating cells with drugs, measuring cell viability, and assessing assay reproducibility (see comparative table below). Based on the data provided in these two studies, it is not possible to determine which experimental procedure (CCLE or GCP) provides more accurate estimates of chemo-sensitivity, as there is no gold-standard set of phenotype measurements to use for comparison and benchmarking. Several published studies and reviews have assessed relative strengths and weaknesses of experimental approaches for assessing chemo-sensitivity. Most of the literature has been focused on assays for measuring cell viability with relatively little published data on systematic comparisons of the methods for earlier steps in the protocols (e.g., media for growing cells, methods for storing compounds, procedures for treating cells)[22–27,29–31,33–38,40–42]. Each of these components may influence drug sensitivity results, and it would be ideal to standardize these steps, where possible.

Perhaps the most significant protocol differences between CCLE and CGP relates to the method of assessing cell viability. CCLE estimated cell viability by bioluminescent quantitation of intracellular ATP content. This well-established method enables assessment of medium and long-term cytotoxic effects and is rapid and extremely sensitive with a large dynamic range[22,28–30,34,36–39]. Limitations of the intracellular ATP assay include the fact that it is unable to identify cell death modes, it is unable to differentiate between lethal and non-lethal perturbations (e.g. contact inhibition, senescence, starvation) producing decreased concentrations of ATP, it can be highly sensitive to metabolic interference, and it is prone to underestimating the efficacy of DNA synthesis-targeting agents[26,29,37,38]. The CGP protocol used a cell-permeant red fluorescent nucleic acid stain (SYTO 60), which releases red fluorescence when binding to nucleic acid from live cells. In contrast to other SYTO probes, the SYTO 60 is unable to distinguish between live cells and cells undergoing early apoptosis[42]. Limitations of this assay include the inability to identify cell death modes[38]. A recent study by Chan and colleagues directly compared cell viability assays based on quantifying total amount of nucleic acid using fluorescent DNA-binding dyes (similar to the SYTO 60 assay used in the CGP study) vs. ATP-dependent luminescence (similar to the assay used in the CCLE study)[26]. The study shows that the ATP-dependent luminescence assay is prone to underestimation of drug potency and efficacy, which was particularly problematic for assessing efficacy of DNA synthesis-targeting agents[26]. The ATP-dependent luminescence and fluorescent DNA-binding assays are measuring different aspects of the drug response phenotype, and therefore it is not surprising that the assays show only moderate correlation in the CGP/CCLE analysis. Given the limitations of each assay, it has been suggested that multi-parameter testing, incorporating multiple, complementary cell-viability assays yields the most robust and informative phenotypic measures[32,33].

An additional area of protocol development and standardization that would likely aid in obtaining robust estimates of chemo-sensitivity would be a more thorough use of controls. CGP used the proteasome inhibitor MG132 as a control, as MG132 is known to be extremely cytotoxic. CCLE used drug-free positive controls and cell-free negative controls. While these controls may establish a bare minimum level of assay function, they are likely inadequate for ensuring accurate quantitative cell viability measurements. Development and distribution of a library of high-quality benchmarked drug-cell line control pairs and associated measurements, ranging from highly sensitive to highly resistant, would likely be useful for ensuring adequate assay function and for estimating accuracy and variability of measurements, compared with a gold-standard set of measurements. Similarly, more systematic use of technical replicates and reporting of raw data values would facilitate statistical estimates of assay reproducibility, which would enable modeling of experimental reproducibility in downstream analyses.

Taken together, the findings from our study and from the prior literature suggest that each of the components reported in the comparative table below can potentially have an important impact

on chemo-sensitivity assays. Further, it is important to note that this is not the only source of experimental error, and even when the exact same protocols are used by multiple investigators (as in the case of Camptothecin in CGP), only fair correlation is obtained between different participating sites. This suggests that additional protocol and method development (beyond standardization) will be important for developing robust and informative chemo-sensitivity measurements.

In summary:

- It is important to attempt to standardize methods for growing cells, storing compounds, treating cells with drugs, measuring cell viability, and assessing assay reproducibility
- It will likely be useful to develop and validate new multi-parametric measures of drug response that are more informative and robust than currently used single parameter approaches The development of more standardized, robust and informative chemo-sensitivity assay procedures will be an important and necessary pre-requisite to enable the application of computational methods to build biologically informative and clinically useful molecular predictors of drug response from large scale pharmacogenomic datasets.

Comparative table

CCLE	CGP
Growth medium	
All cell lines were cultured in RPMI or DMEM with 10% fetal bovine serum (FBS; Invitrogen)	Cells were grown in RPMI or DMEM/F12 medium supplemented with 5% FBS and penicillin/streptavidin
Exclude cross contamination and synonymous lines	
SNP fingerprint using Affymetrix SNP array 6.0 (20,000 randomly selected SNPs)	SNP fingerprinting using Sequenom (92 SNPs) and short tandem repeat (STR) analysis using AmpFlSTRIdentifiler, Applied Biosystems
Optimal cell number measurement	
Not specified	70% cell confluency/ensure reaching growth phase
Storage of compounds	
Compounds were dissolved in 90% DMSO/10% water at 2 mM and stored at -20°C until use	Compounds were stored as 10?mM aliquots at -80°C, and were subjected to a maximum of five freeze-thaw cycles
Plating Cells	
Cell lines were dispensed into 1,536-well plates (optimized for tissue culture) with a final volume of 5 μL and a concentration of 250 cells per well	Cells were seeded in either 96-well or 384-well microplates

CCLE	CGP
Drug concentration range	
Drugs serially diluted, concentration range of 2 mM to 636 nM	The range of concentrations selected for each compound was based on in vitro data of concentrations inhibiting relevant kinase activity and cell viability
Colony formation assays	
Not specified	Yes
Adherent Cells	
12 to 24 hours after plating, 20 nL of each compound dilution-cell mix, were incubated for 72 to 84 hours	Adherent cell lines were plated 1 day before treatment with a 9-point twofold dilution series of compounds and assayed at a 72 hours time point
Suspension Cells	
Not specified	Suspension cell lines were treated with compound immediately following plating, incubated for 72 hours, and then stained with 55μg ml-1 resazurin (Sigma)
Viability assay	
Cell numbers were determined by measuring the amount of ATP per well using Cell Titer Glo (Promega)	Cells stained with 1μM of the fluorescent nucleic acid stain Syto60 (Invitrogen) for 1 hour
Use of controls	
Wells containing vehicle only or the positive control compound MG132 (a proteasome inhibitor toxic to most cell lines at 1μM) were also included	Sixteen (96-well format) or 42 (384-well) drug-free positive controls, 8 (96-well) or 32 (384-well) negative (no cells) controls
Assays reproducibility	
Compounds were tested in duplicate, occasionally, lines were assayed multiple times (weeks to months apart, data not shown)	Drug screening was performed at two sites using matched cell line collections (data available for Camptothecin, drugs number 1003 and 195)

3.1 GSK study

The GSK authors used the same pharmacological assay used by the CCLE (Cell Titer Glo Luminescent Cell Viability Assay kit from Promega), but other parameters in the experimental protocols differ from those in either CGP or CCLE For instance they tested a different range of drug concentrations (0.0003 μM, 0.0032 μM, 0.01 μM, 0.032 μM, 0.1 μM, 0.317 μM, 1 μM, 3.16 μM, and 10 μM) and they used yet another model to estimate IC50 values (model 205 in XLfit in Microsoft Excel).

4 Supplementary Tables

IC50 sensitivity calling
ERLOTINIB

CCLE vs CGP	res	inter	sens
res	43	21	5
inter	1	1	0
sens	0	0	0

Kappa=0.022, 95%CI [−0.17,0.22], p=9.7E−01

IC50 sensitivity calling
LAPATINIB

CCLE vs CGP	res	inter	sens
res	39	13	2
inter	10	10	1
sens	0	0	0

Kappa=0.22, 95%CI [−0.004,0.44], p=4.0E−01

IC50 sensitivity calling
PHA665752

CCLE vs CGP	res	inter	sens
res	25	37	4
inter	5	6	1
sens	0	1	0

Kappa=−0.017, 95%CI [−0.19,0.16], p=8.5E−01

IC50 sensitivity calling
CRIZOTINIB

CCLE vs CGP	res	inter	sens
res	24	23	1
inter	15	10	5
sens	0	1	1

Kappa=−0.03, 95%CI [−0.24,0.18], p=4.0E−02

IC50 sensitivity calling
TAE684

CCLE vs CGP	res	inter	sens
res	12	23	9
inter	9	14	10
sens	0	1	2

Kappa=−0.00024, 95%CI [−0.19,0.19], p=3.7E−01

IC50 sensitivity calling
NILOTINIB

CCLE vs CGP	res	inter	sens
res	112	25	0
inter	37	10	1
sens	1	0	2

Kappa=0.082, 95%CI [−0.069,0.23], p=1.3E−03

IC50 sensitivity calling
AZD0530

CCLE vs CGP	res	inter	sens
res	13	39	16
inter	0	3	6
sens	0	0	1

Kappa=0.0085, 95%CI [−0.11,0.13], p=3.1E−02

IC50 sensitivity calling
SORAFENIB

CCLE vs CGP	res	inter	sens
res	15	29	5
inter	4	10	11
sens	0	0	2

Kappa=0.029, 95%CI [−0.15,0.21], p=2.1E−03

IC50 sensitivity calling
PD0332991

CCLE vs CGP	res	inter	sens
res	49	74	33
inter	4	12	11
sens	0	0	2

Kappa=0.041, 95%CI [−0.055,0.14], p=2.2E−02

IC50 sensitivity calling
PLX4720

CCLE vs CGP	res	inter	sens
res	148	54	3
inter	8	14	7
sens	2	0	2

Kappa=0.21, 95%CI [0.087,0.33], p=6.2E−08

IC50 sensitivity calling
PD0325901

CCLE vs CGP	res	inter	sens
res	53	85	29
inter	3	13	25
sens	2	6	19

Kappa=0.1, 95%CI [0.012,0.19], p=1.1E−10

IC50 sensitivity calling
AZD6244

CCLE vs CGP	res	inter	sens
res	48	100	37
inter	4	5	20
sens	1	2	5

Kappa=−0.015, 95%CI [−0.1,0.072], p=2.1E−06

IC50 sensitivity calling
NUTLIN3

CCLE vs CGP	res	inter	sens
res	33	132	58
inter	0	3	9
sens	0	1	1

Kappa=−0.0076, 95%CI [−0.065,0.05], p=7.4E−03

IC50 sensitivity calling
17AAG

CCLE vs CGP	res	inter	sens
res	32	14	1
inter	25	107	37
sens	0	11	15

Kappa=0.34, 95%CI [0.23,0.45], p=3.3E−16

IC50 sensitivity calling
PACLITAXEL

CCLE vs CGP	res	inter	sens
res	7	6	5
inter	11	14	9
sens	5	9	10

Kappa=0.1, 95%CI [−0.099,0.3], p=6.4E−01

Supplementary Table 1 Contingency tables comparing the sensitivity calls (res, inter, and sens standing for resistant, intermediate and sensitive drug phenotype, respectively) computed from IC_{50} measures for each of the 15 drugs screened both in CGP and CCLE. The Kappa coefficient, its confidence interval and its significance are reported below each contingency table.

AUC sensitivity calling ERLOTINIB

CCLE vs CGP	res	inter	sens
res	49	5	6
inter	5	1	3
sens	1	0	1

Kappa=0.15, 95%CI [−0.072,0.37], p=2.9E−01

AUC sensitivity calling LAPATINIB

CCLE vs CGP	res	inter	sens
res	56	5	4
inter	4	1	1
sens	1	1	2

Kappa=0.25, 95%CI [−0.0031,0.49], p=8.3E−02

AUC sensitivity calling PHA665752

CCLE vs CGP	res	inter	sens
res	57	5	3
inter	5	1	1
sens	5	0	2

Kappa=0.17, 95%CI [−0.08,0.41], p=2.5E−01

AUC sensitivity calling CRIZOTINIB

CCLE vs CGP	res	inter	sens
res	53	4	1
inter	11	1	2
sens	3	1	4

Kappa=0.26, 95%CI [0.042,0.48], p=3.1E−03

AUC sensitivity calling TAE684

CCLE vs CGP	res	inter	sens
res	51	11	2
inter	6	2	1
sens	5	0	2

Kappa=0.12, 95%CI [−0.11,0.35], p=1.3E−01

AUC sensitivity calling NILOTINIB

CCLE vs CGP	res	inter	sens
res	162	10	6
inter	5	1	0
sens	1	0	3

Kappa=0.23, 95%CI [0.0044,0.45], p=2.7E−03

AUC sensitivity calling AZD0530

CCLE vs CGP	res	inter	sens
res	61	5	2
inter	3	3	1
sens	0	0	3

Kappa=0.48, 95%CI [0.23,0.74], p=5.0E−05

AUC sensitivity calling SORAFENIB

CCLE vs CGP	res	inter	sens
res	51	1	5
inter	10	1	4
sens	0	3	1

Kappa=0.2, 95%CI [0.017,0.38], p=1.8E−04

AUC sensitivity calling PD0332991

CCLE vs CGP	res	inter	sens
res	43	21	22
inter	11	4	16
sens	27	7	34

Kappa=0.1, 95%CI [−0.027,0.23], p=7.3E−03

AUC sensitivity calling PLX4720

CCLE vs CGP	res	inter	sens
res	174	10	9
inter	18	2	1
sens	11	4	9

Kappa=0.24, 95%CI [0.1,0.38], p=2.6E−05

AUC sensitivity calling PD0325901

CCLE vs CGP	res	inter	sens
res	92	6	0
inter	28	6	2
sens	48	28	25

Kappa=0.24, 95%CI [0.15,0.33], p=1.6E−13

AUC sensitivity calling AZD6244

CCLE vs CGP	res	inter	sens
res	143	11	2
inter	17	8	9
sens	12	3	17

Kappa=0.42, 95%CI [0.31,0.54], p=2.1E−15

AUC sensitivity calling NUTLIN3

CCLE vs CGP	res	inter	sens
res	139	14	40
inter	16	3	9
sens	7	1	8

Kappa=0.12, 95%CI [0.0067,0.24], p=9.3E−02

AUC sensitivity calling 17AAG

CCLE vs CGP	res	inter	sens
res	54	7	9
inter	29	25	60
sens	6	11	41

Kappa=0.28, 95%CI [0.19,0.37], p=3.3E−16

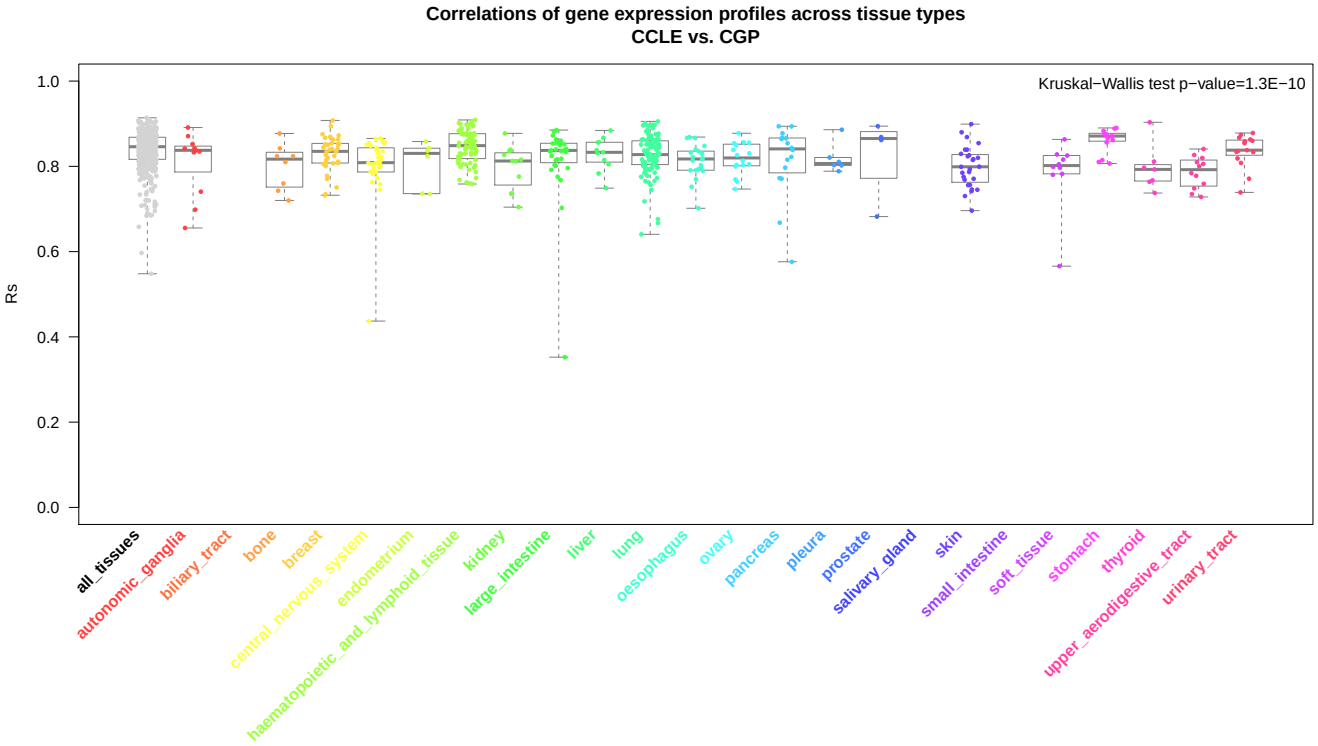
AUC sensitivity calling PACLITAXEL

CCLE vs CGP	res	inter	sens
res	12	4	5
inter	18	4	20
sens	2	0	11

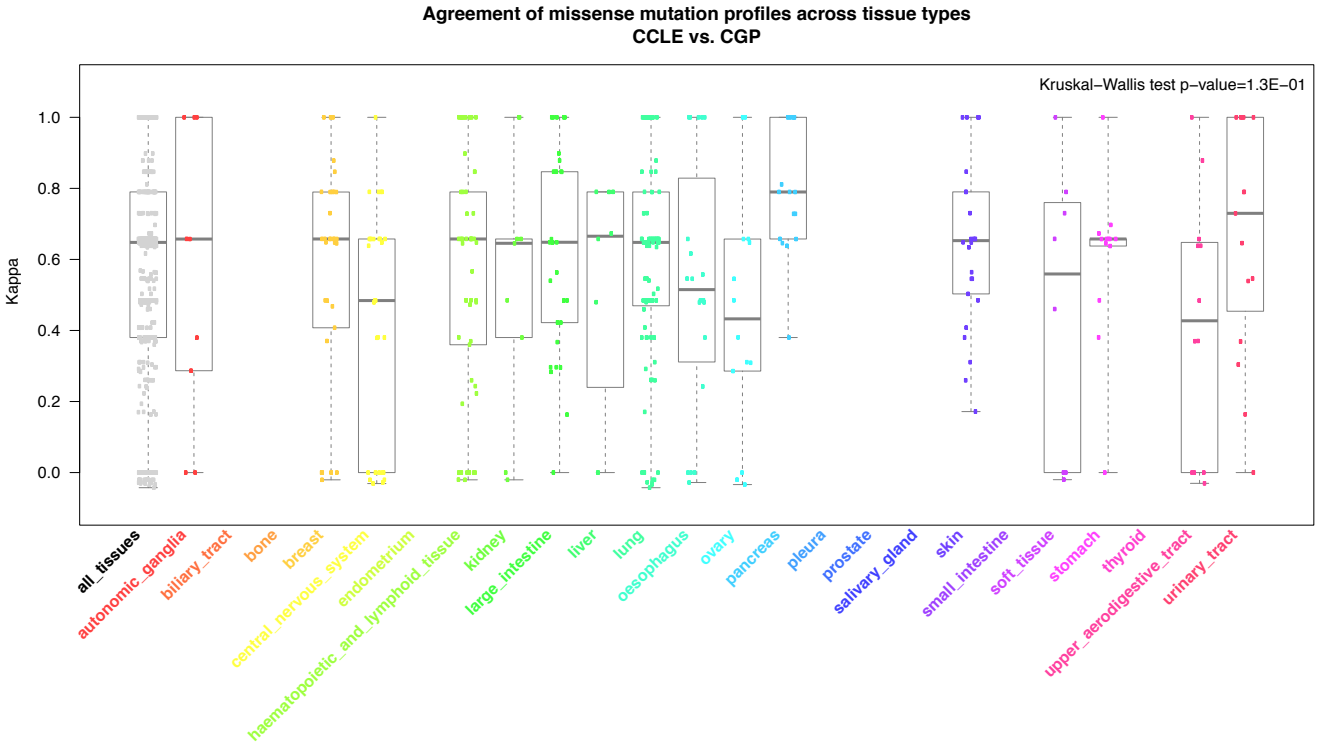
Kappa=0.13, 95%CI [−0.02,0.29], p=7.4E−03

Supplementary Table 2 Contingency tables comparing the sensitivity calls (res, inter, and sens standing for resistant, intermediate and sensitive drug phenotype, respectively) computed from AUC measures for each of the 15 drugs screened both in CGP and CCLE. The Kappa coefficient, its confidence interval and its significance are reported below each contingency table.

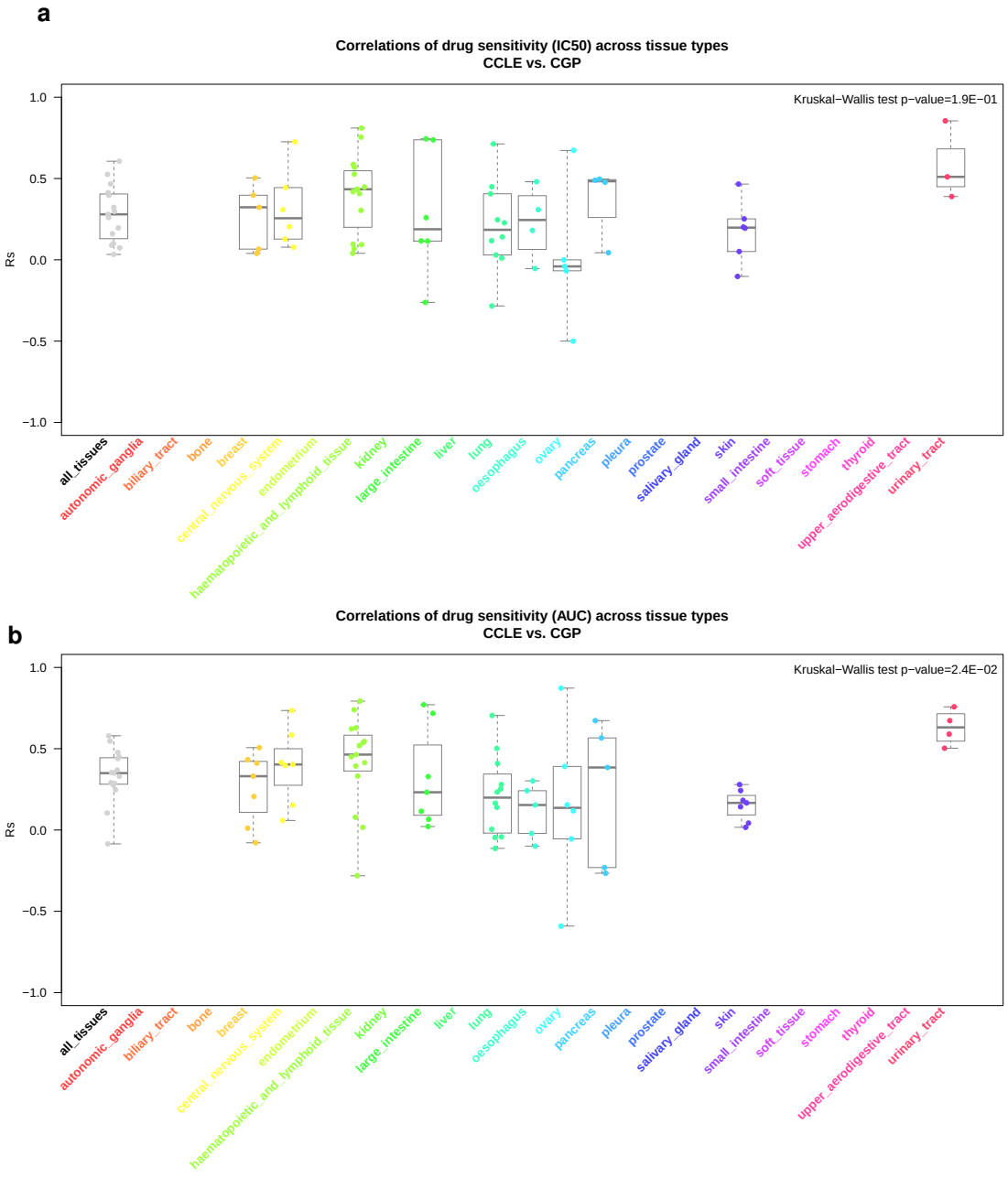
5 Supplementary Figures



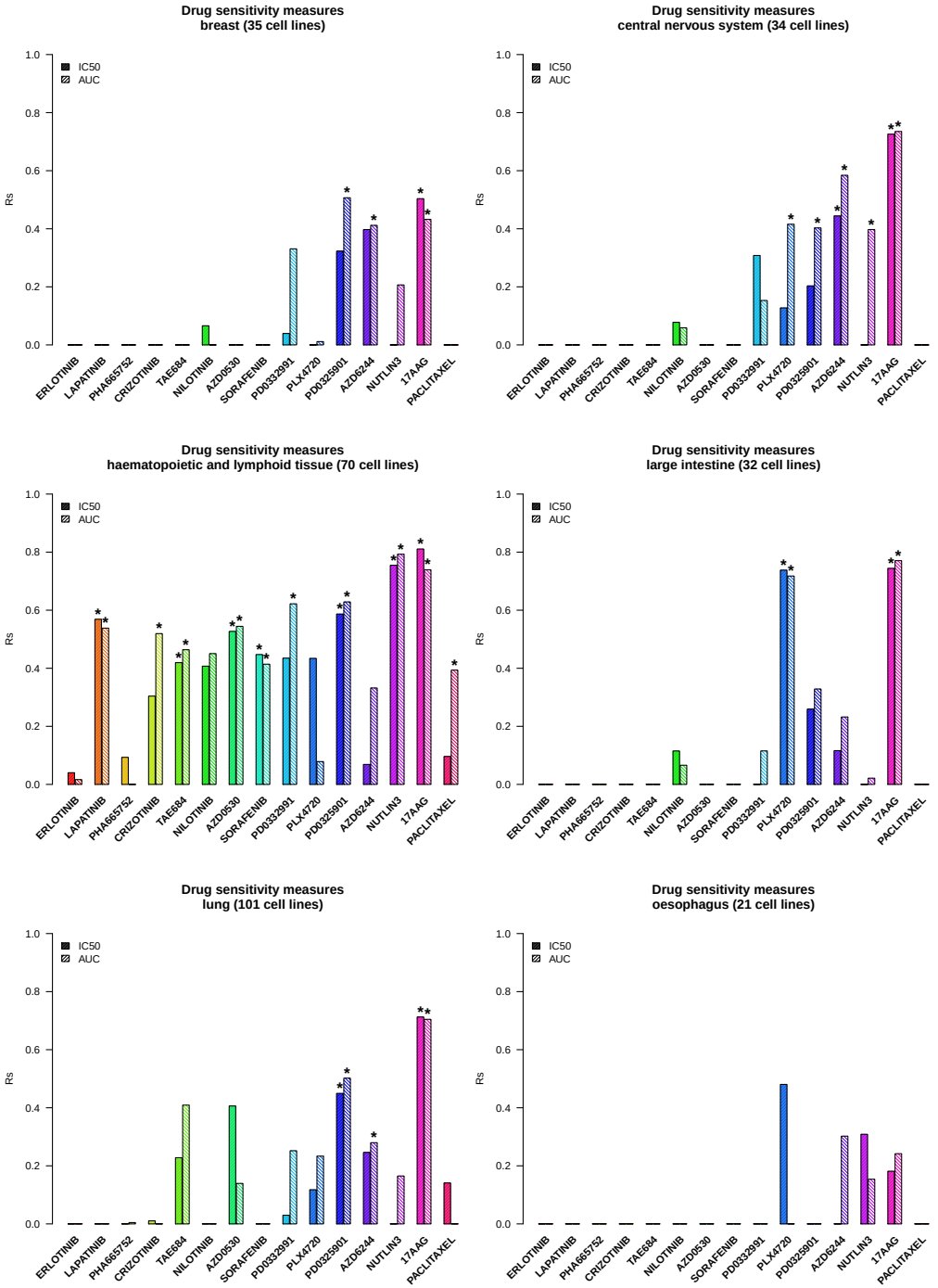
Supplementary Figure 1 Box plot of the correlations of gene expression profiles between identical cell lines in CGP and CCLE, across tissue types. Kruskal-Wallis test was used to test whether correlations significantly depended on tissue type (upper right corner).



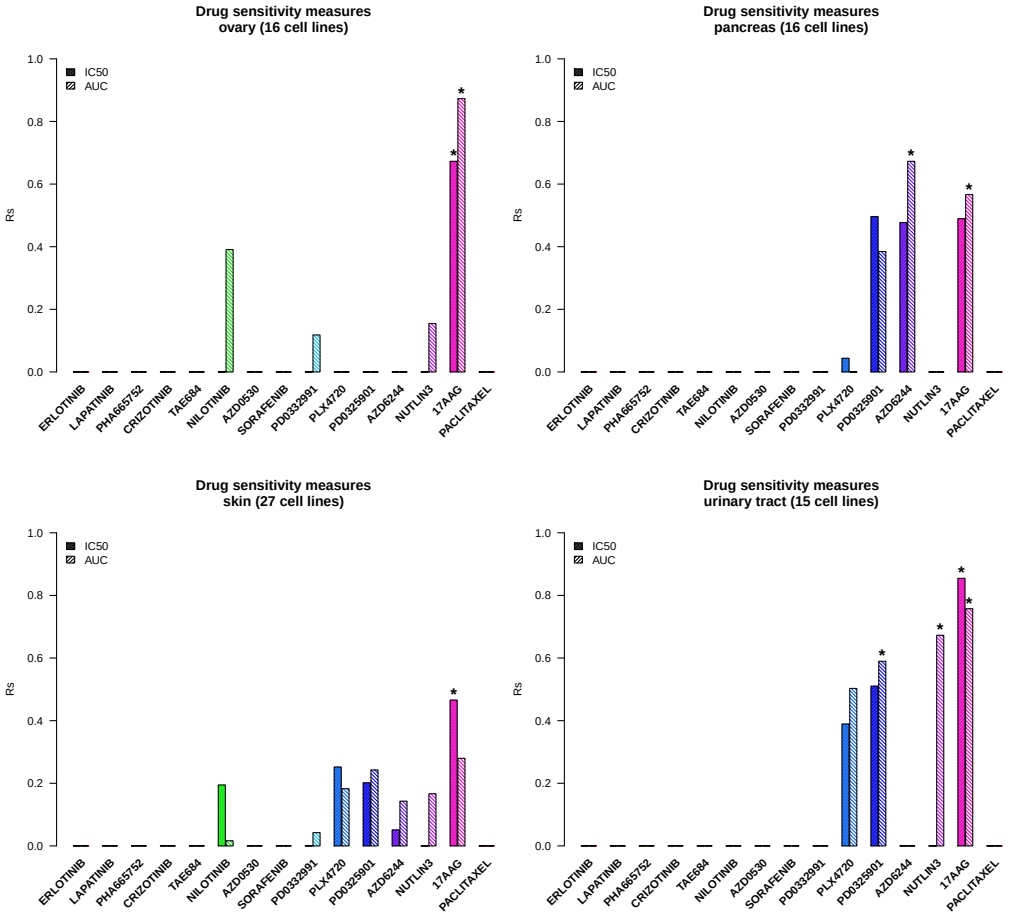
Supplementary Figure 2 Box plot of the correlations of (missense) mutation profiles between identical cell lines in CGP and CCLE, across tissue types. Kruskal-Wallis test was used to test whether agreement significantly depended on tissue type (upper right corner).



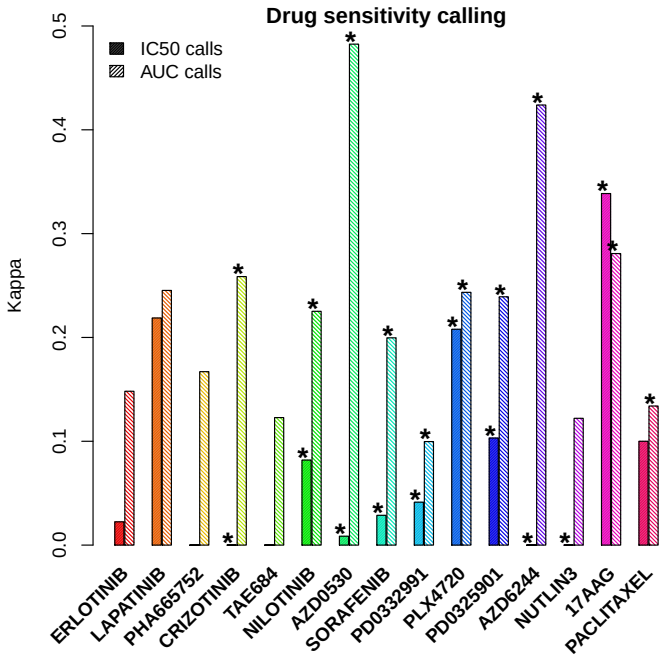
Supplementary Figure 3 Box plot of the correlations of the sensitivity measures for 15 drugs, across tissue types. (a) Correlations between IC_{50} measures; (b) correlations between AUC measures. Correlations were estimated using the Spearman coefficient (R_s). Kruskal-Wallis test was used to test whether correlations significantly depended on tissue type (upper right corner).



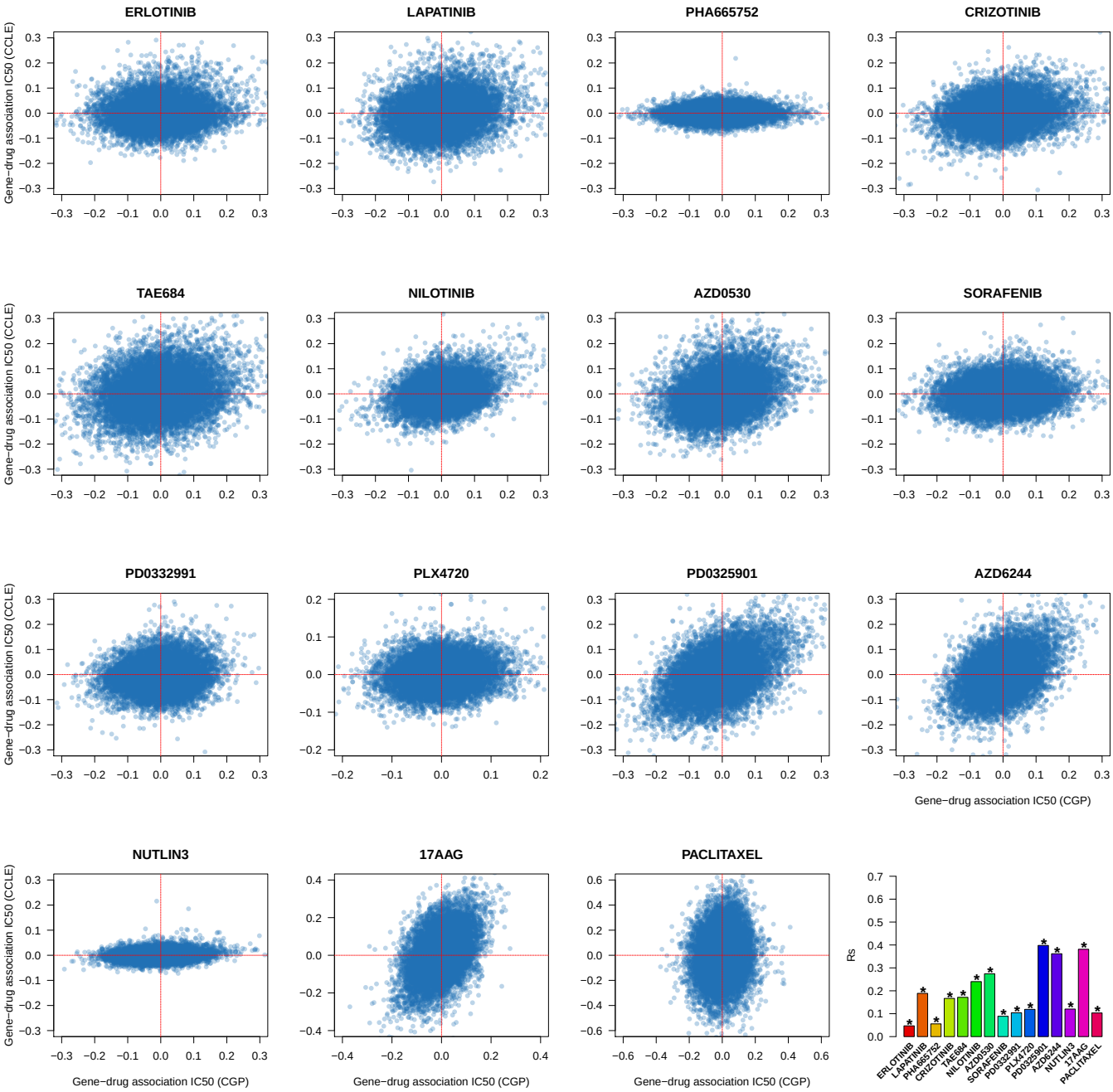
Supplementary Figure 4 Bar plot reporting the positive Spearman correlation coefficients (R_s) for drug sensitivity computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value $<$ 0.05.



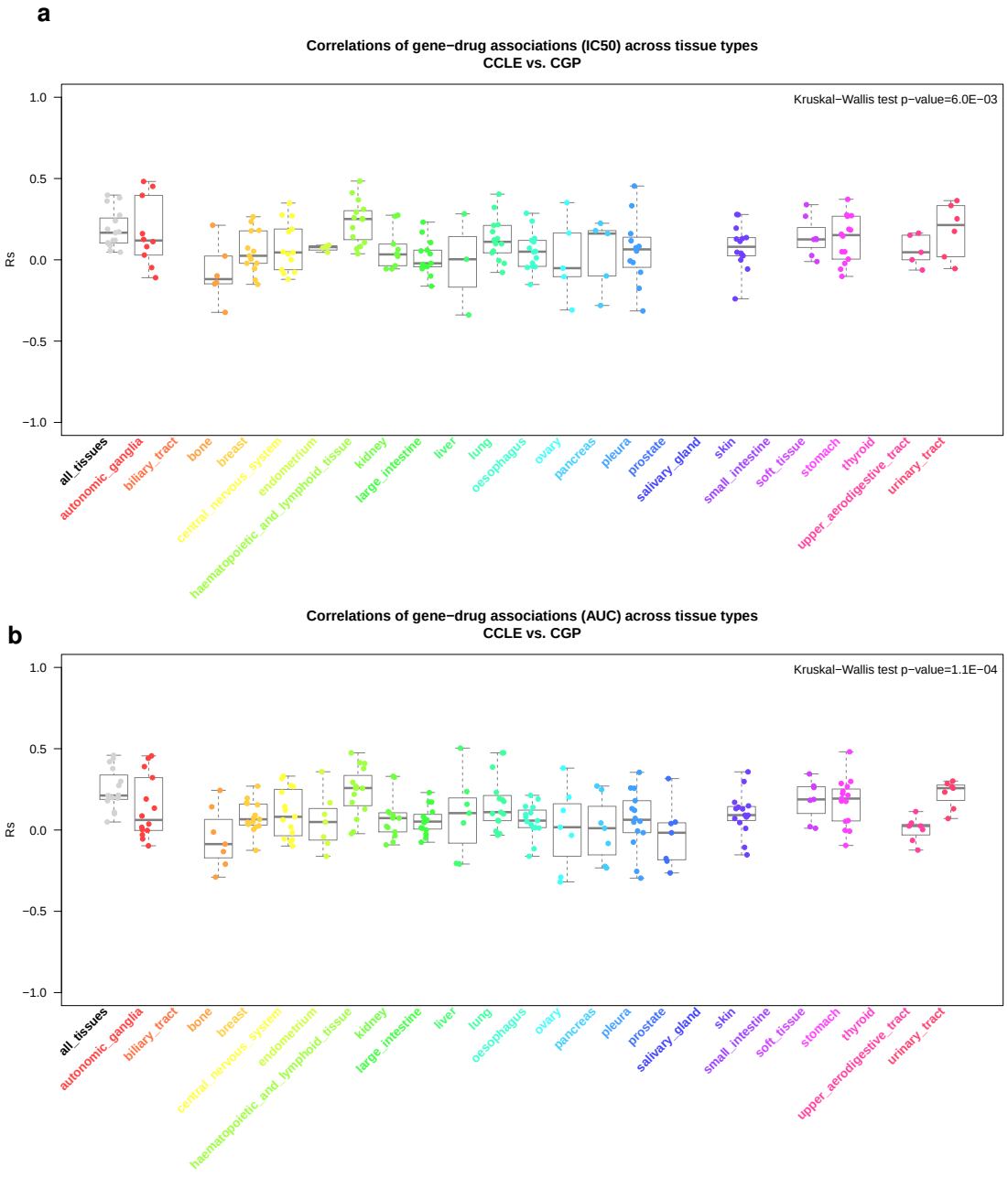
Supplementary Figure 4 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for drug sensitivity computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05.



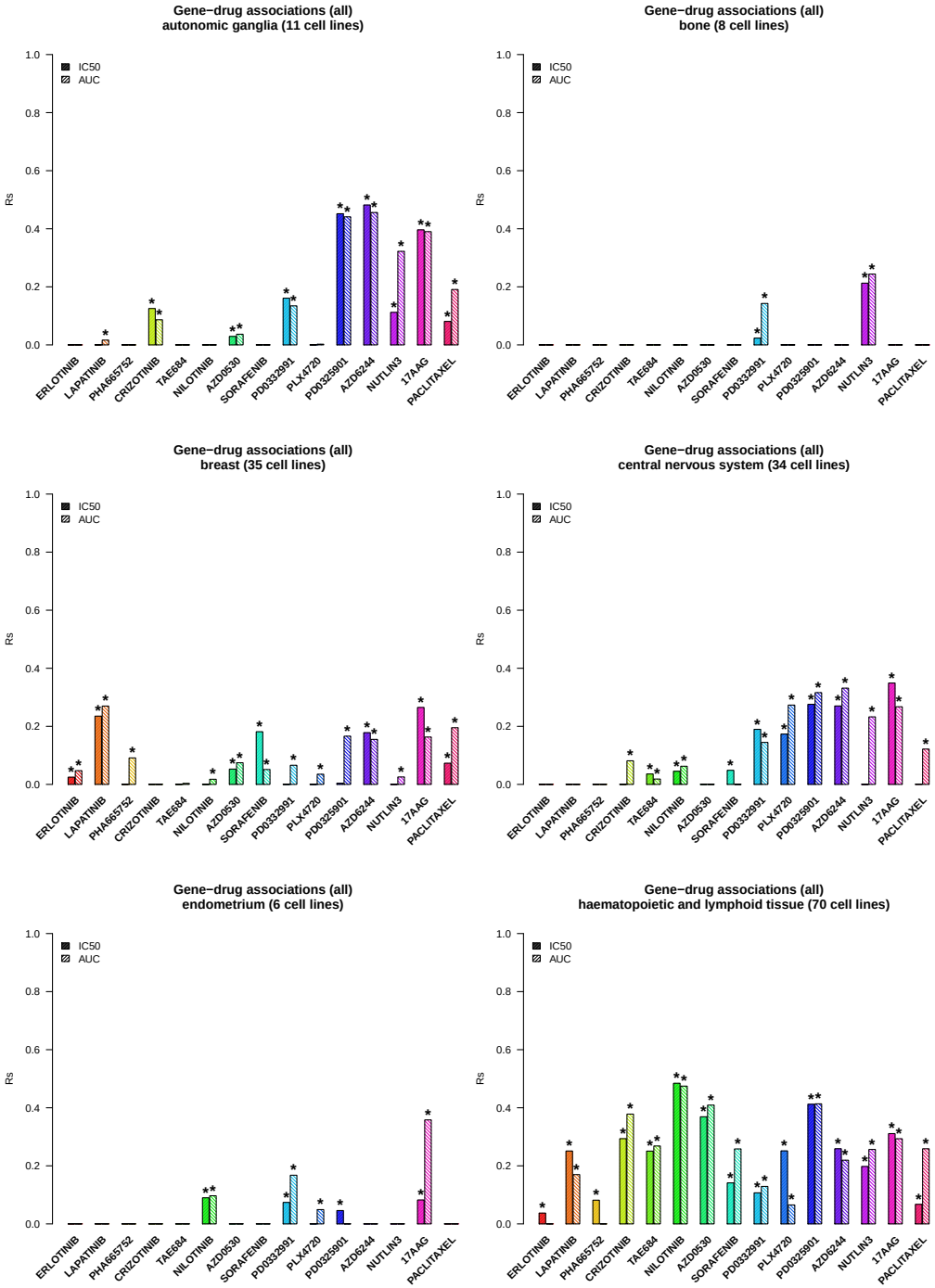
Supplementary Figure 5 Bar plot reporting Cohen's Kappa coefficients (K) quantitatively assessing the concordance between drug sensitivity calls computed with IC_{50} and AUC measures both in CGP and CCLE.



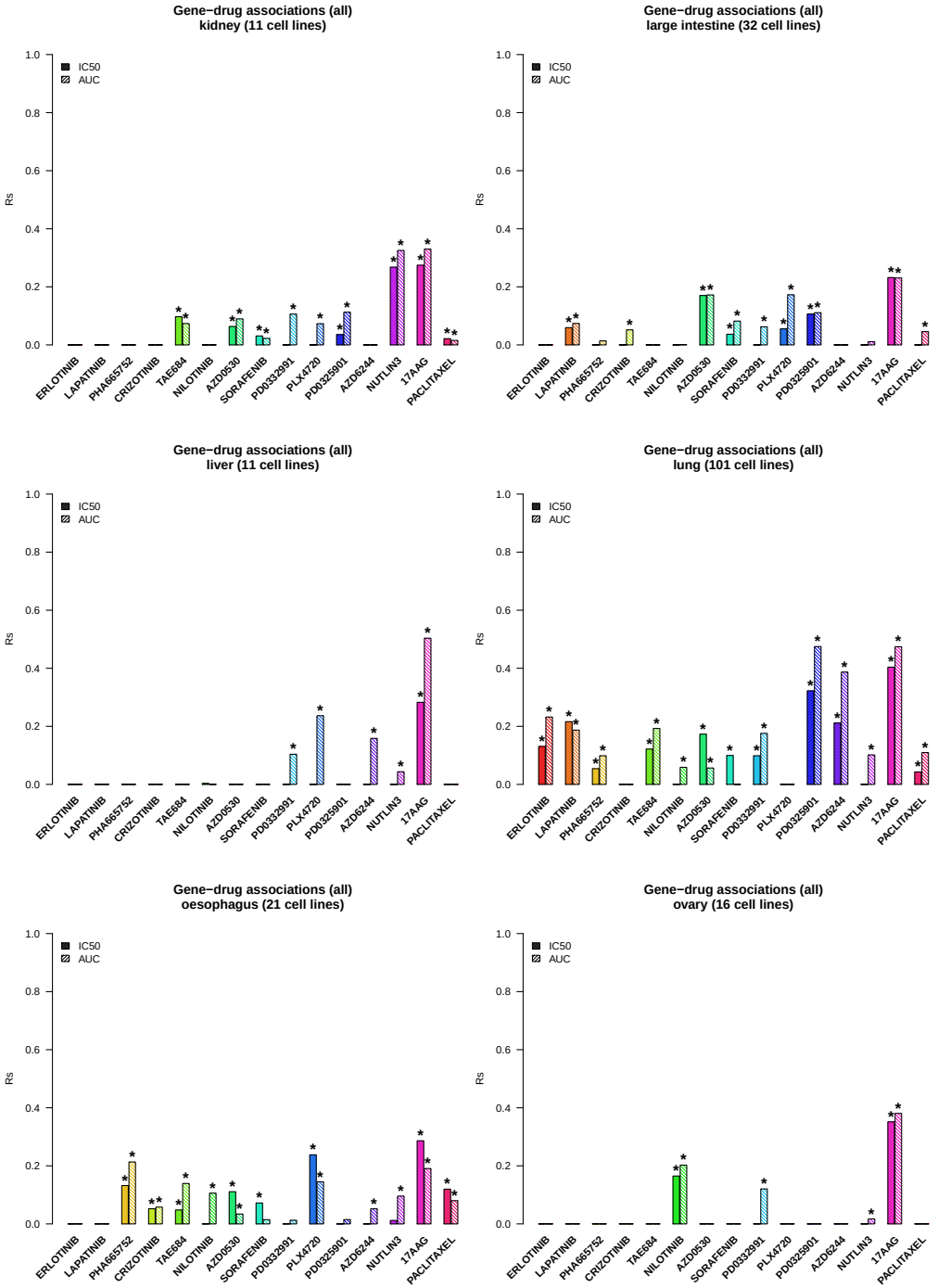
Supplementary Figure 6 Scatter plots reporting the gene-drug associations computed with IC_{50}, as quantified by the standardized coefficient of the gene of interest in a linear model controlled for tissue type, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE. The last bar plot (bottom right corner) reports the Spearman correlation coefficient (R_s) for each drug.



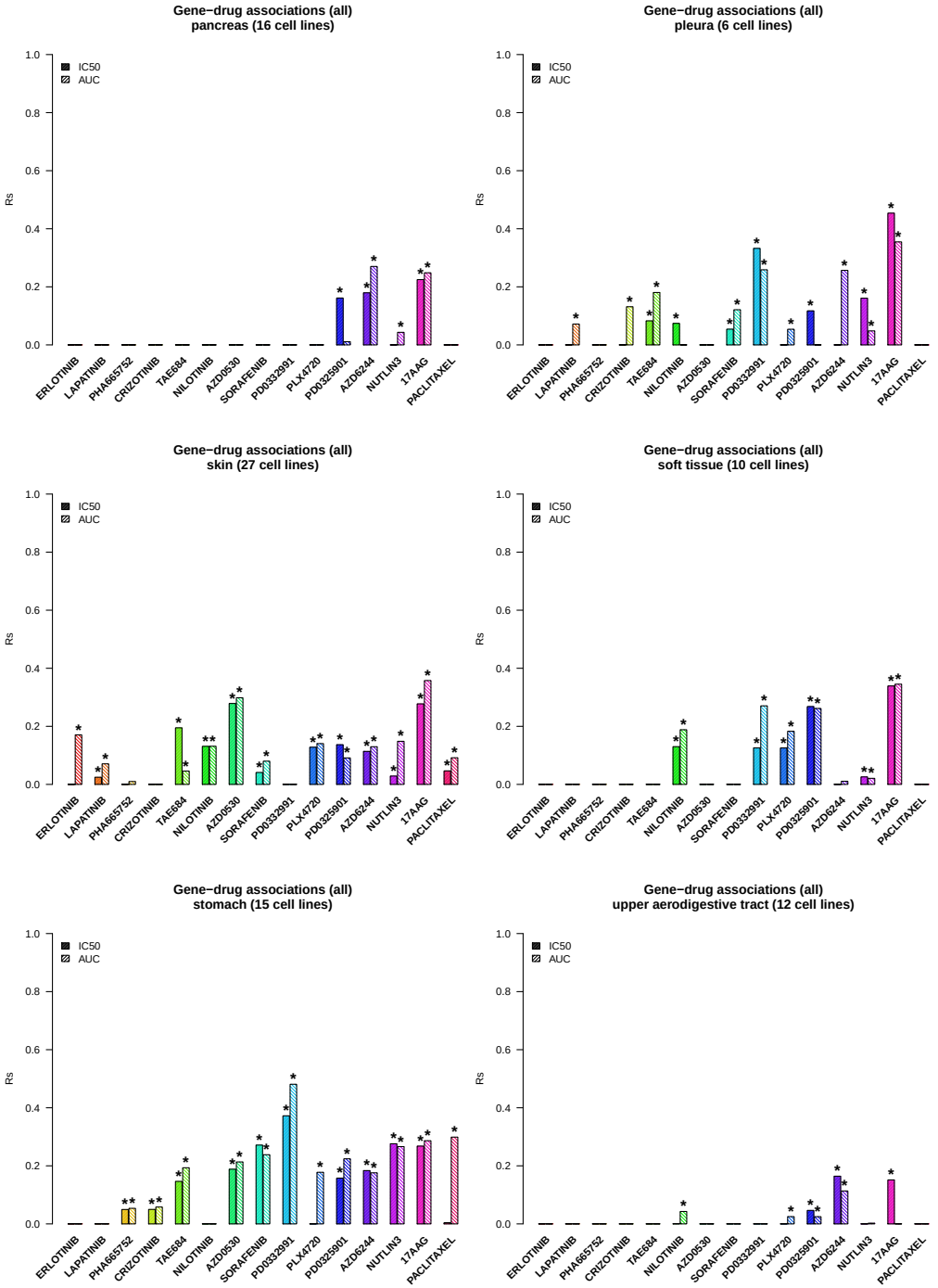
Supplementary Figure 7 Box plot of the correlations of the gene-drug associations for the 15 drugs, across tissue types. (a) Correlations between gene-drug associations computed with IC_{50} in CGP and CCLE; (b) correlations between gene-drug associations computed with AUC in CGP and CCLE. Correlations were estimated using the Spearman coefficient (R_s). Kruskal-Wallis test was used to test whether correlations significantly depended on tissue type (upper right corner).



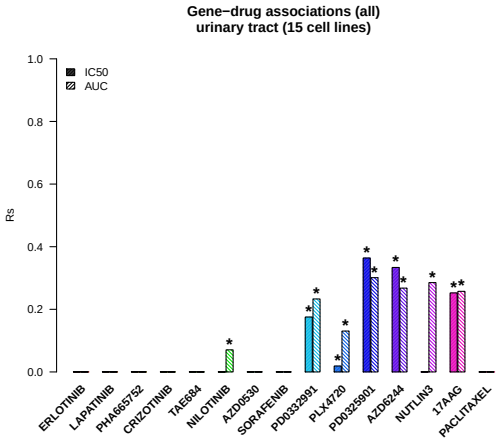
Supplementary Figure 8 Bar plot reporting the positive Spearman correlation coefficients (R_s) for gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



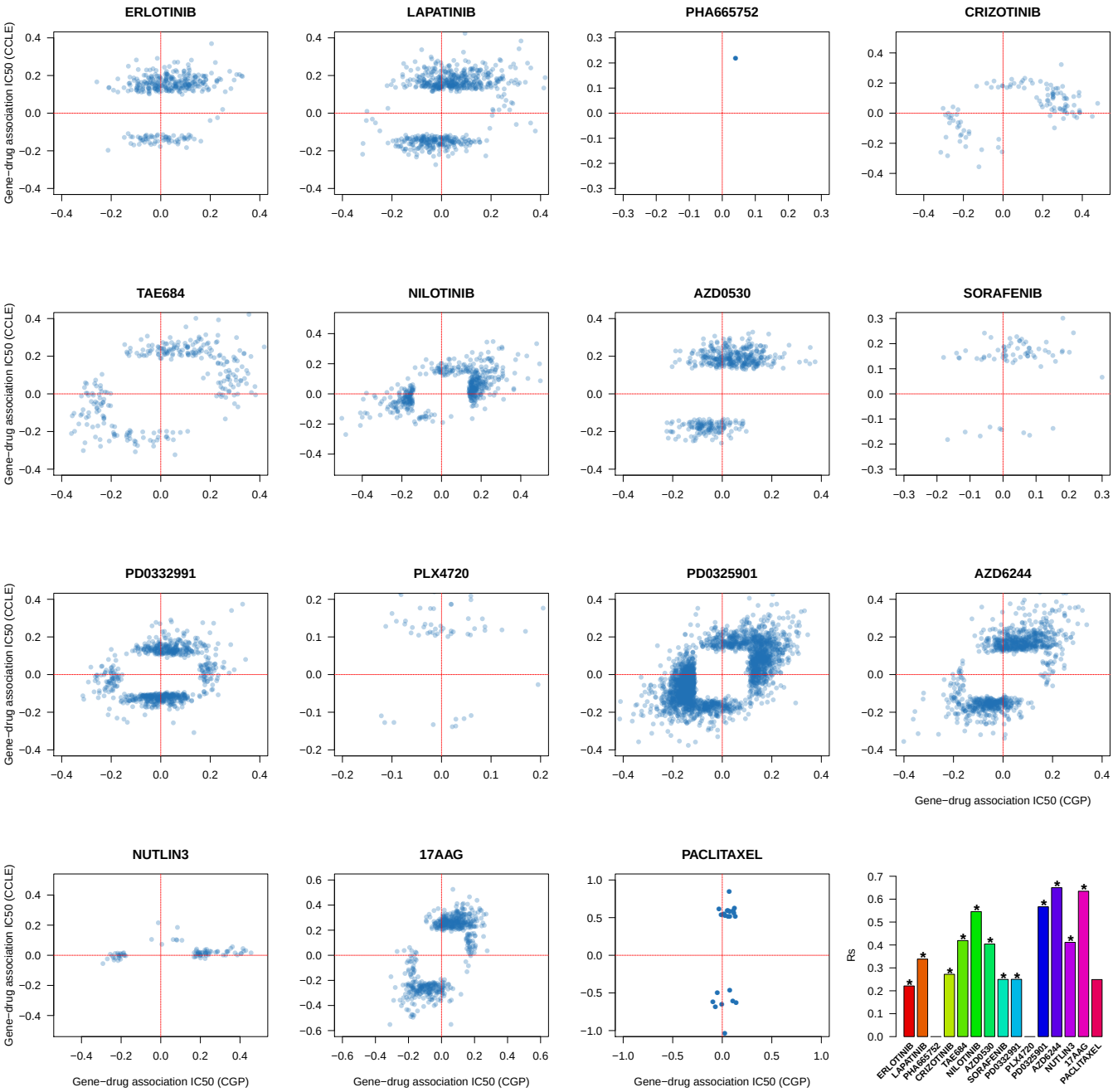
Supplementary Figure 8 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



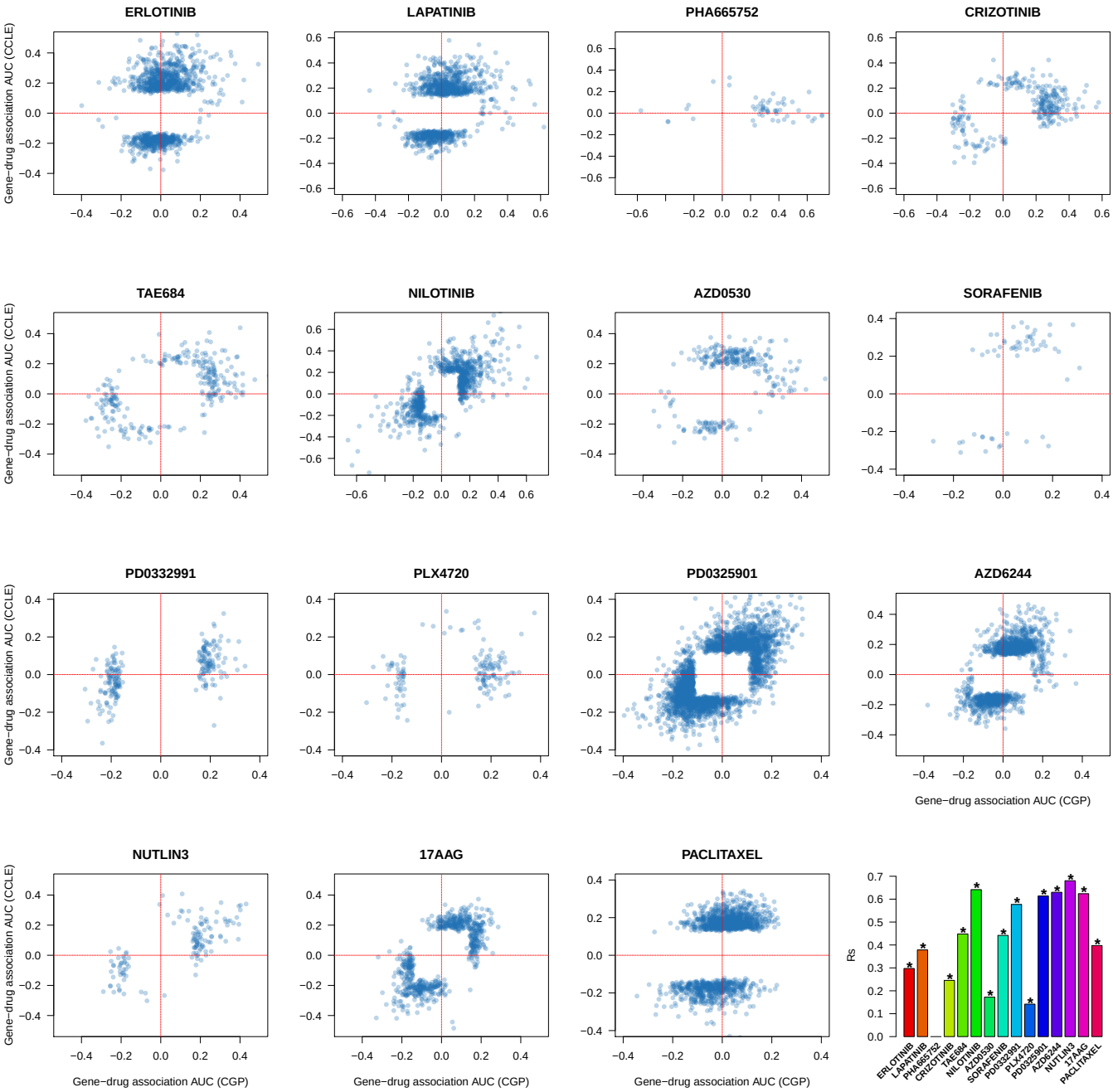
Supplementary Figure 8 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



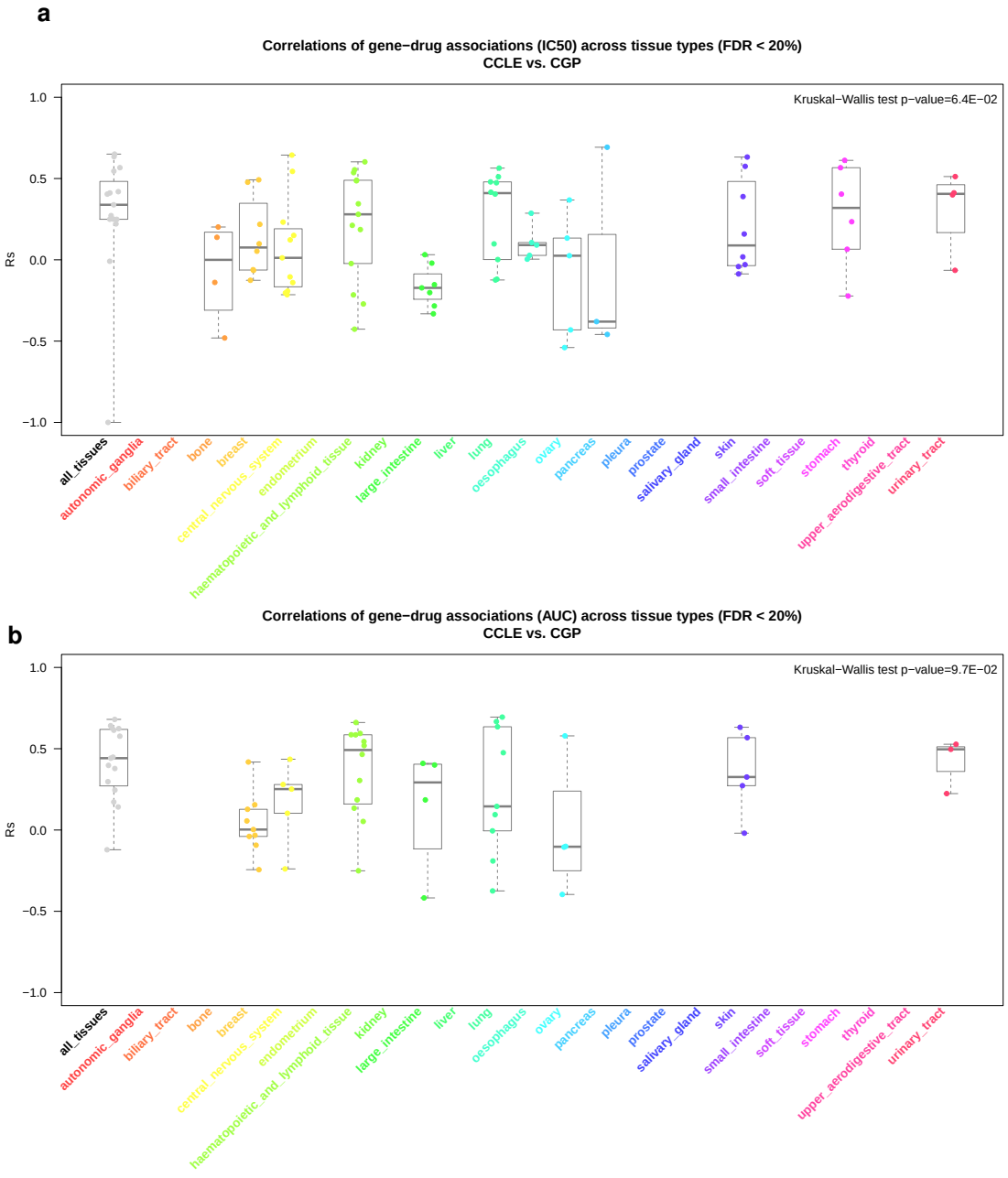
Supplementary Figure 8 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



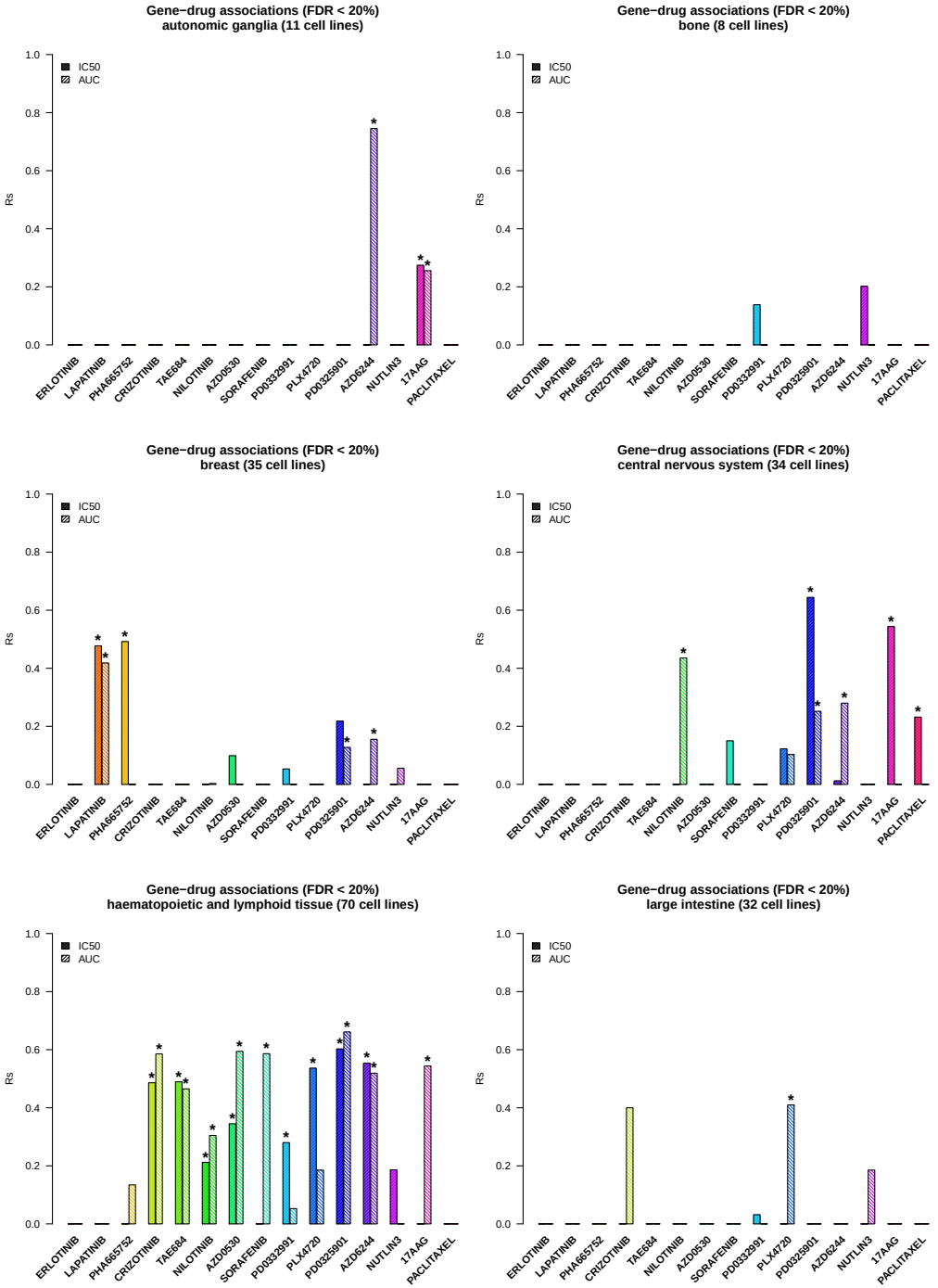
Supplementary Figure 9 Scatter plots reporting the significant (FDR<20%) gene-drug associations computed with IC_{50}, as quantified by the standardized coefficient of the gene of interest in a linear model controlled for tissue type, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE. The last bar plot (bottom right corner) reports the Spearman correlation coefficient (R_s) for each drug.



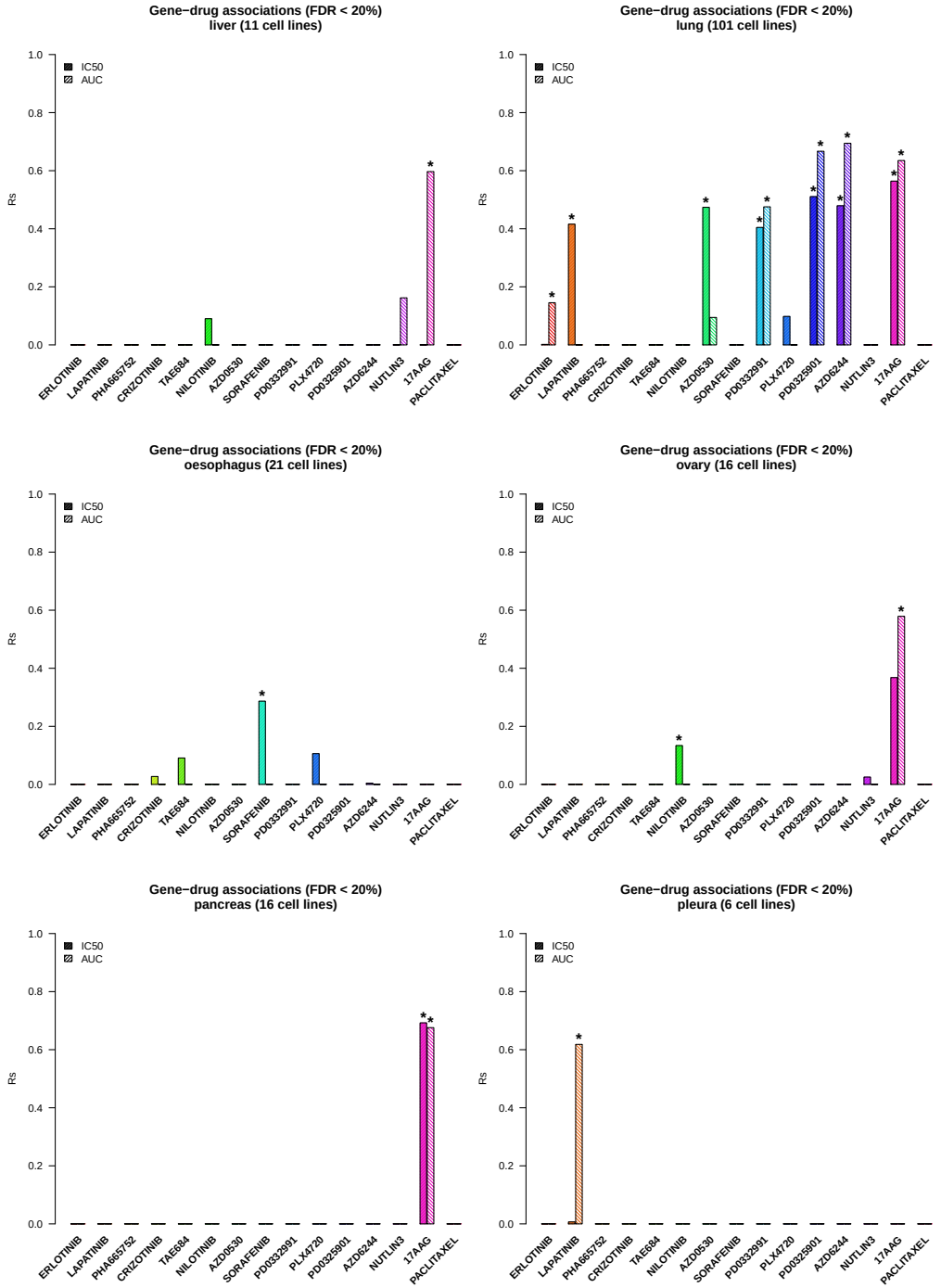
Supplementary Figure 10 Scatter plots reporting the significant (FDR<20%) gene-drug associations computed with AUC, as quantified by the standardized coefficient of the gene of interest in a linear model controlled for tissue type, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE; (b) The last bar plot (bottom right corner) reports the Spearman correlation coefficient (R_s) for each drug.



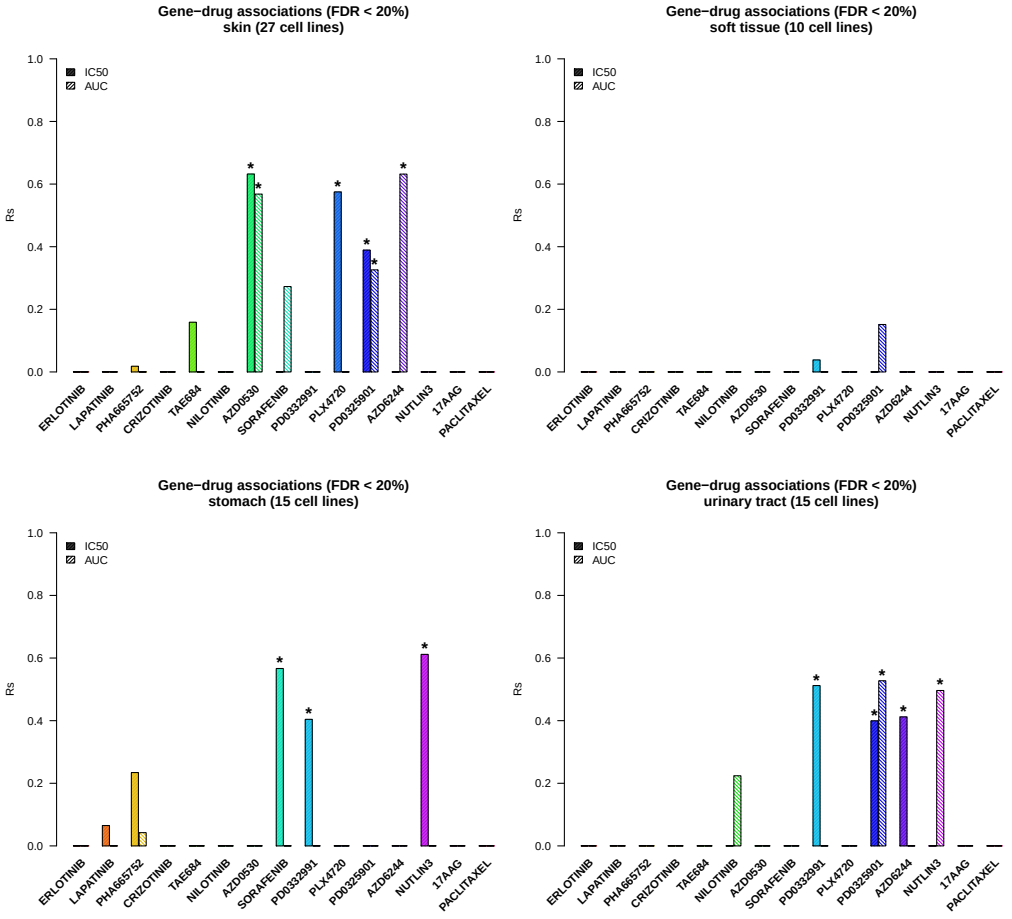
Supplementary Figure 11 Box plot of the correlations of the significant (FDR $<$ 20%) gene-drug associations for the 15 drugs, across tissue types. (a) Correlations between gene-drug associations computed with IC_{50} in CGP and CCLE; (b) correlations between gene-drug associations computed with AUC in CGP and CCLE. Correlations were estimated using the Spearman coefficient (R_s). Kruskal-Wallis test was used to test whether correlations significantly depended on tissue type (upper right corner).



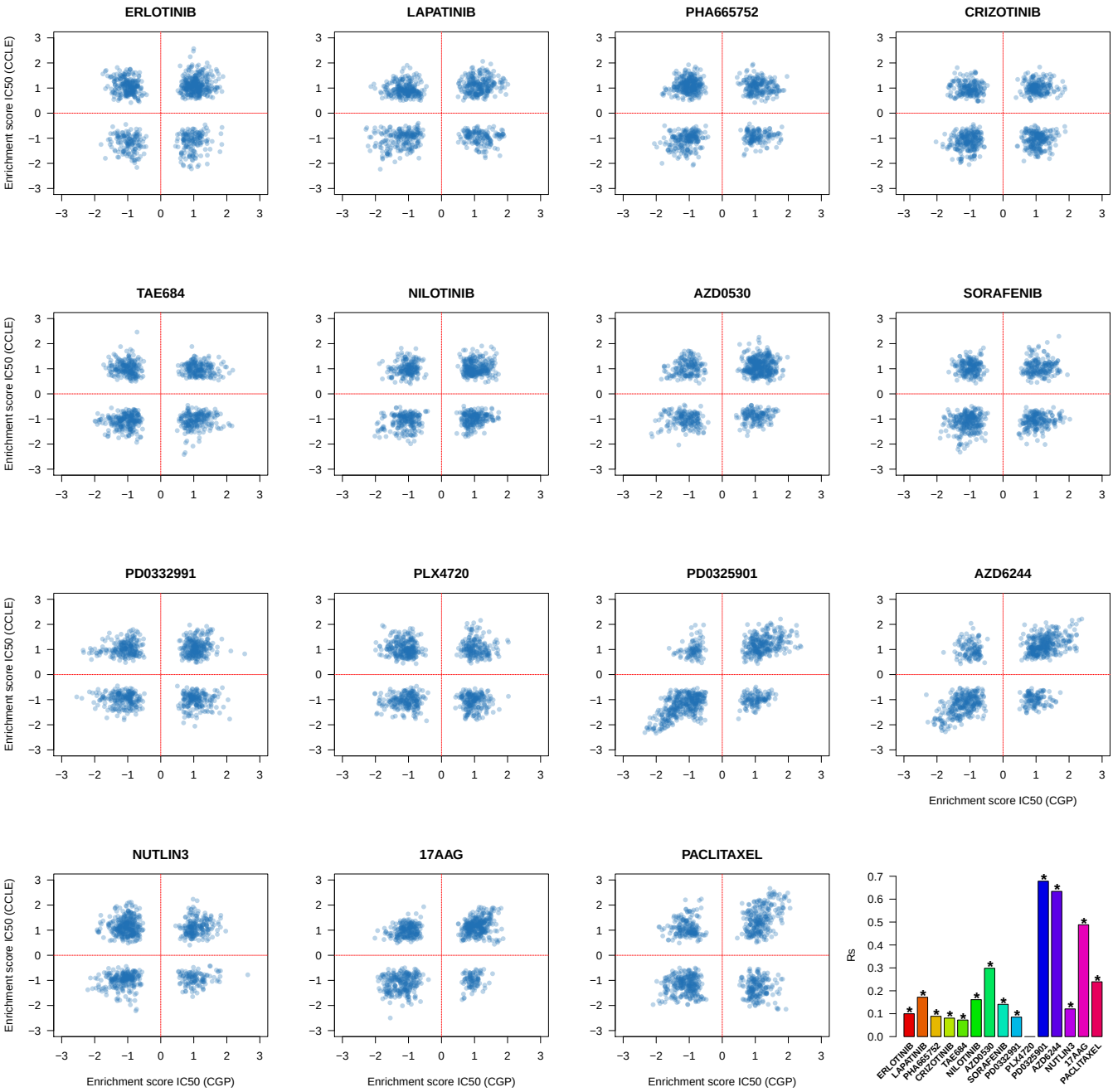
Supplementary Figure 12 Bar plot reporting the positive Spearman correlation coefficients (R_s) for significant (FDR $< 20\%$) gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



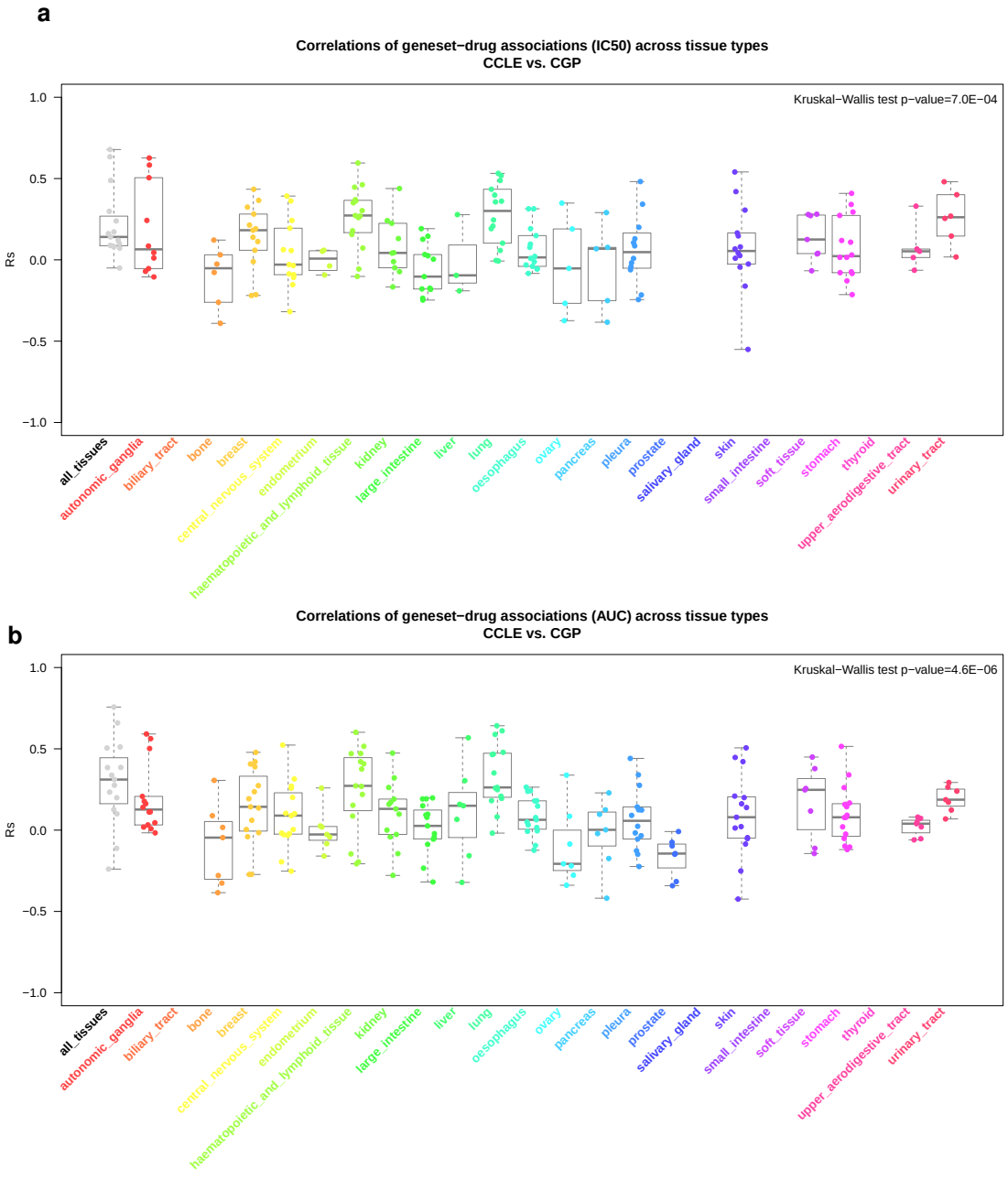
Supplementary Figure 12 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for significant (FDR $<$ 20%) gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value $<$ 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



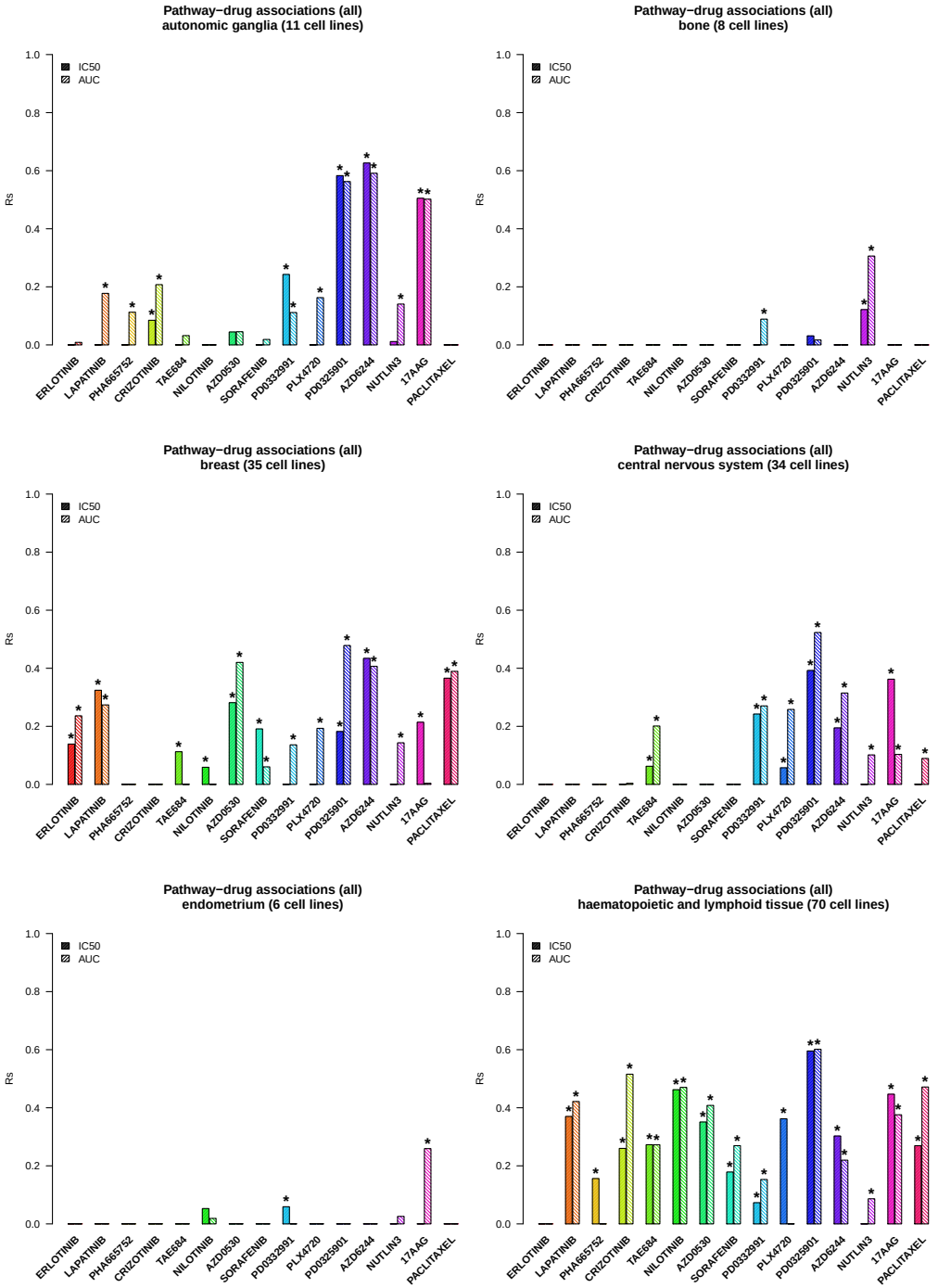
Supplementary Figure 12 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for significant (FDR $<$ 20%) gene-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value $<$ 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



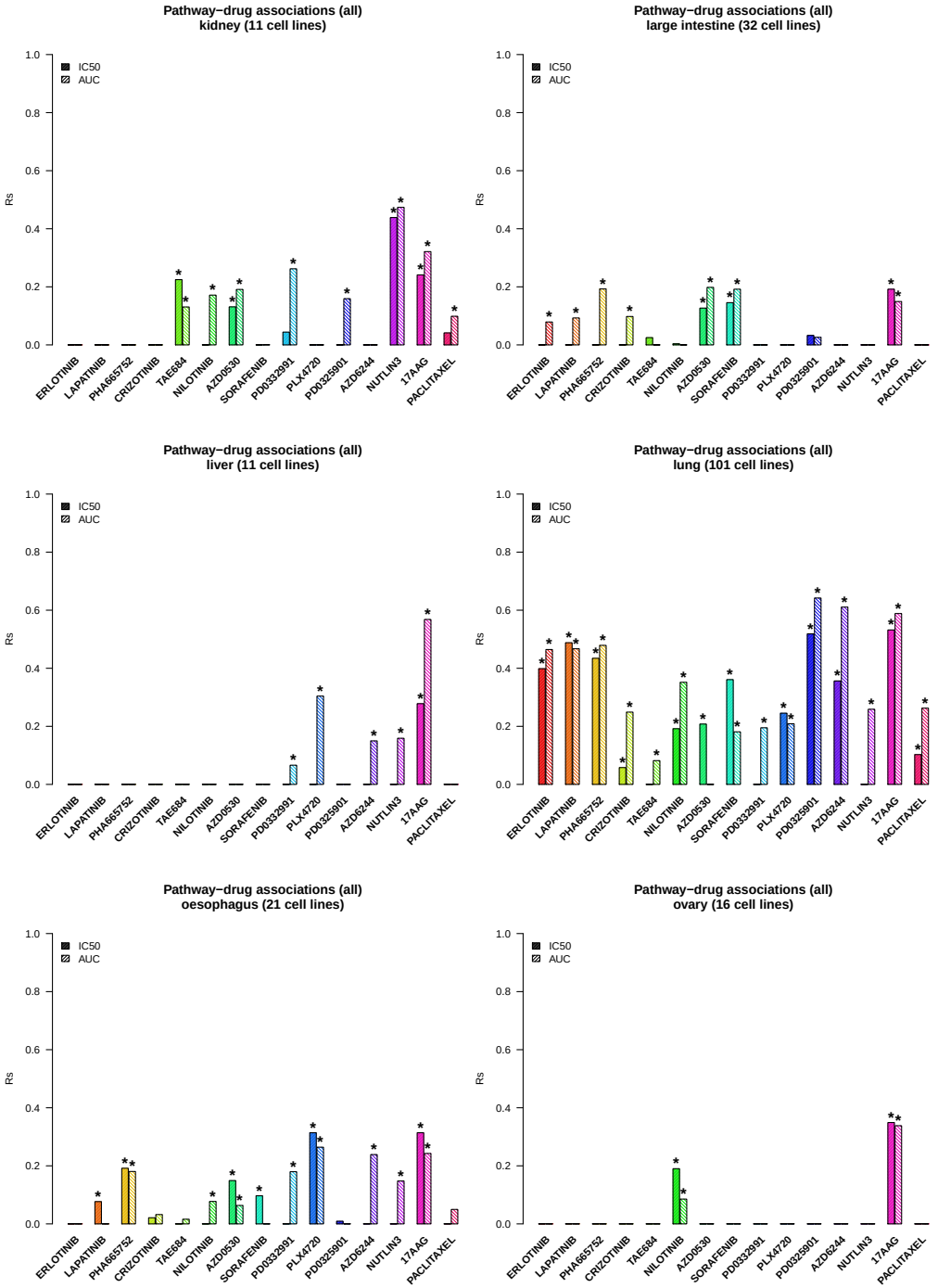
Supplementary Figure 13 Scatter plots reporting the pathway-drug associations computed with IC_{50}, as quantified by the enrichment score from gene set enrichment analysis, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE. The last bar plot (bottom right corner) reports the Spearman correlation coefficient (R_s) for each drug.



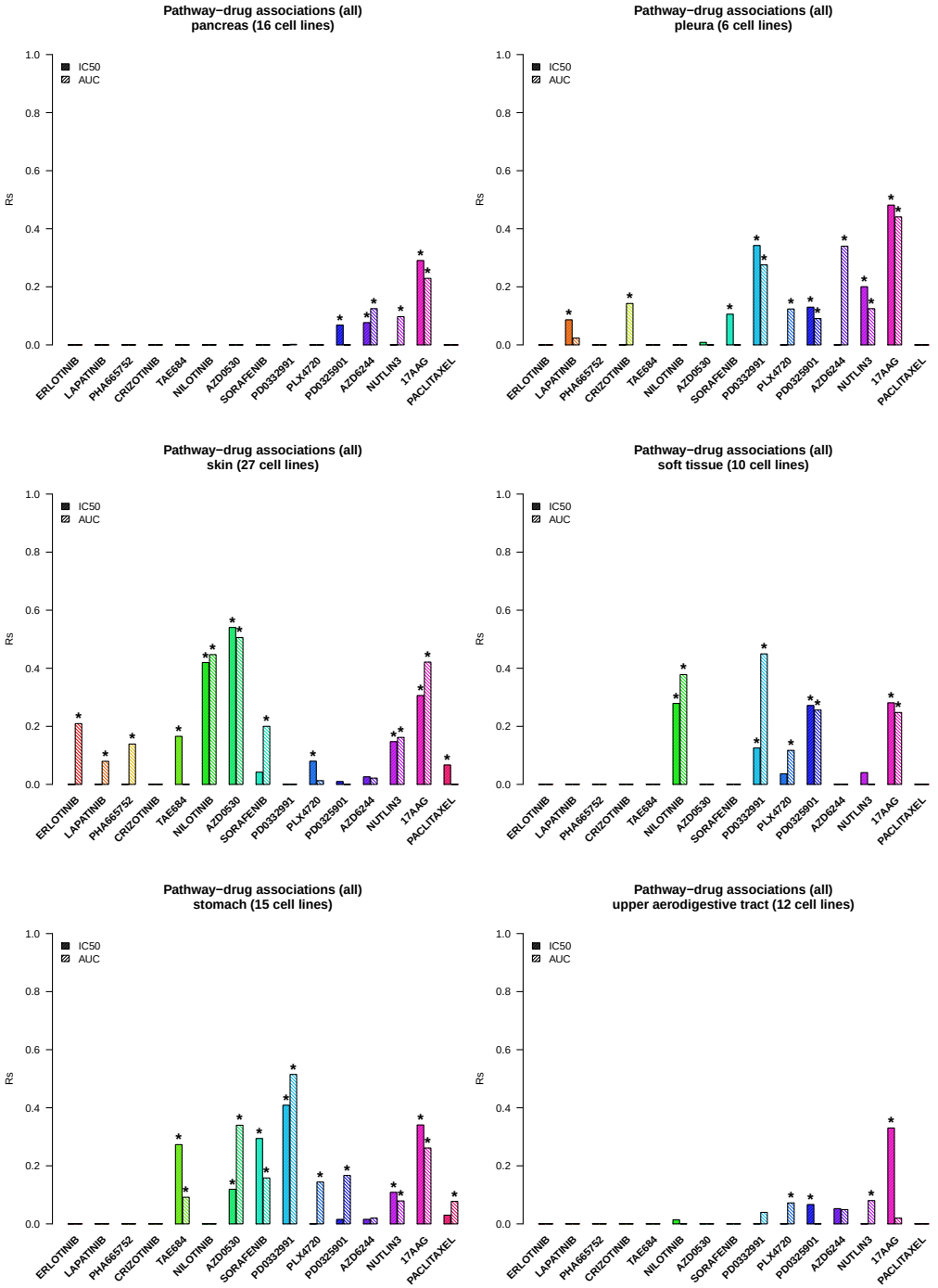
Supplementary Figure 14 Box plot of the correlations of the pathway-drug associations for the 15 drugs, across tissue types. (a) Correlations between pathway-drug associations computed with IC_{50} in CGP and CCLE; (b) correlations between pathway-drug associations computed with AUC in CGP and CCLE. Correlations were estimated using the Spearman coefficient (R_s). Kruskal-Wallis test was used to test whether correlations significantly depended on tissue type (upper right corner).



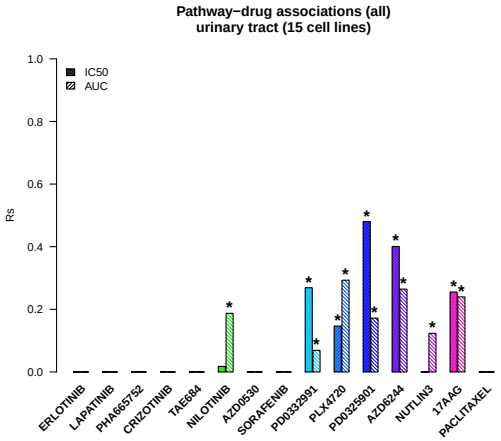
Supplementary Figure 15 Bar plot reporting the positive Spearman correlation coefficients (R_s) for pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



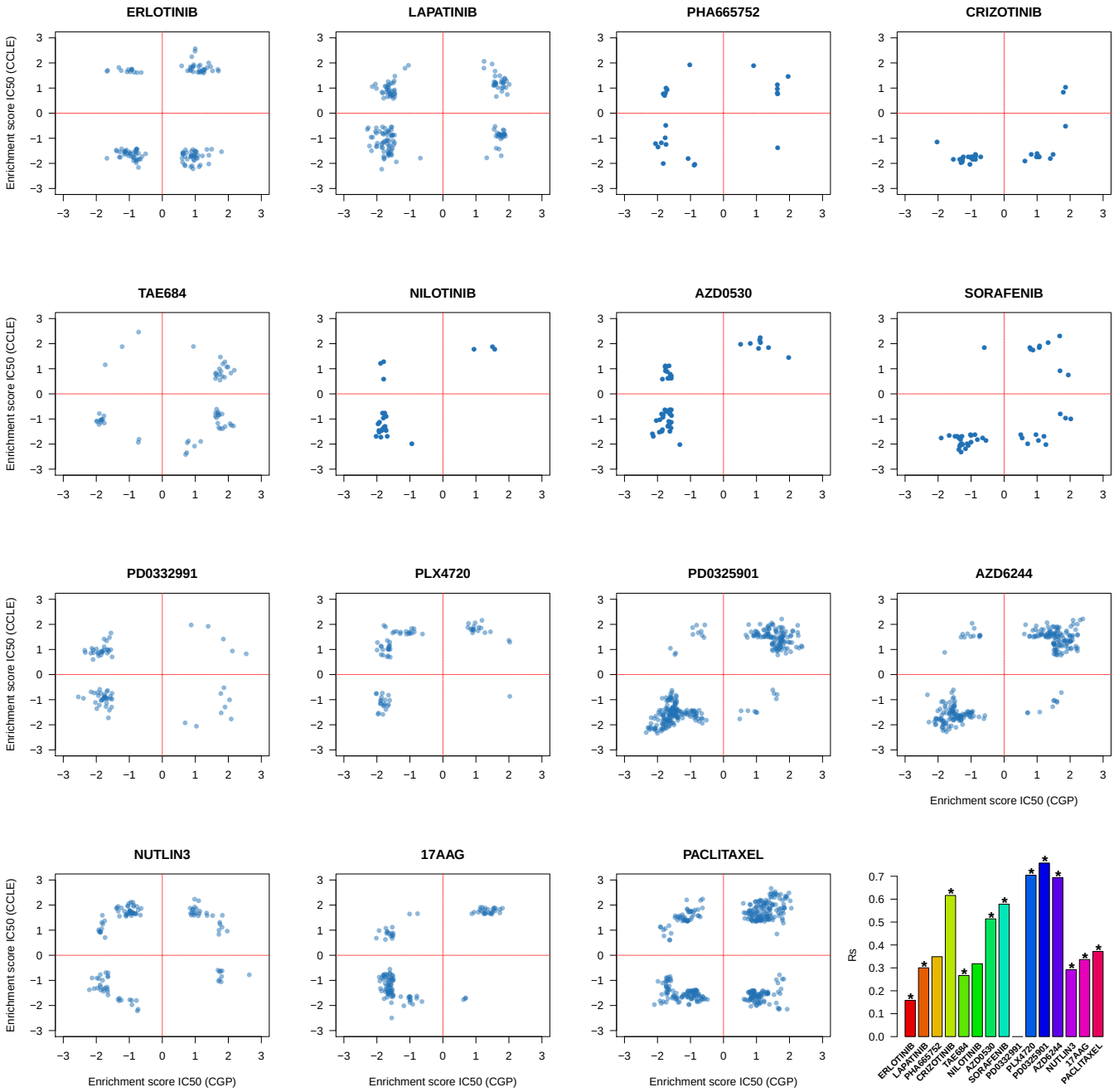
Supplementary Figure 15 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



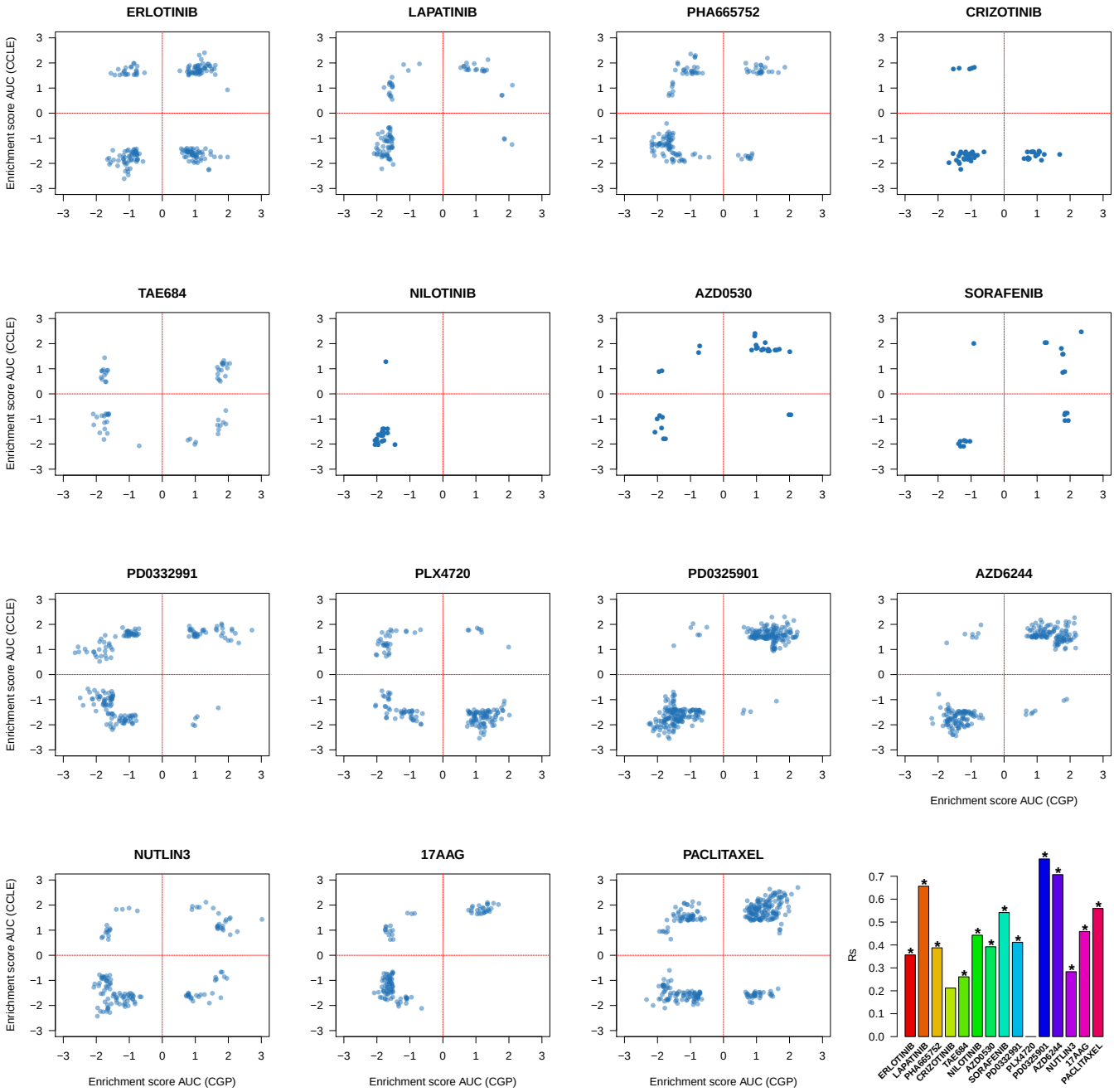
Supplementary Figure 15 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



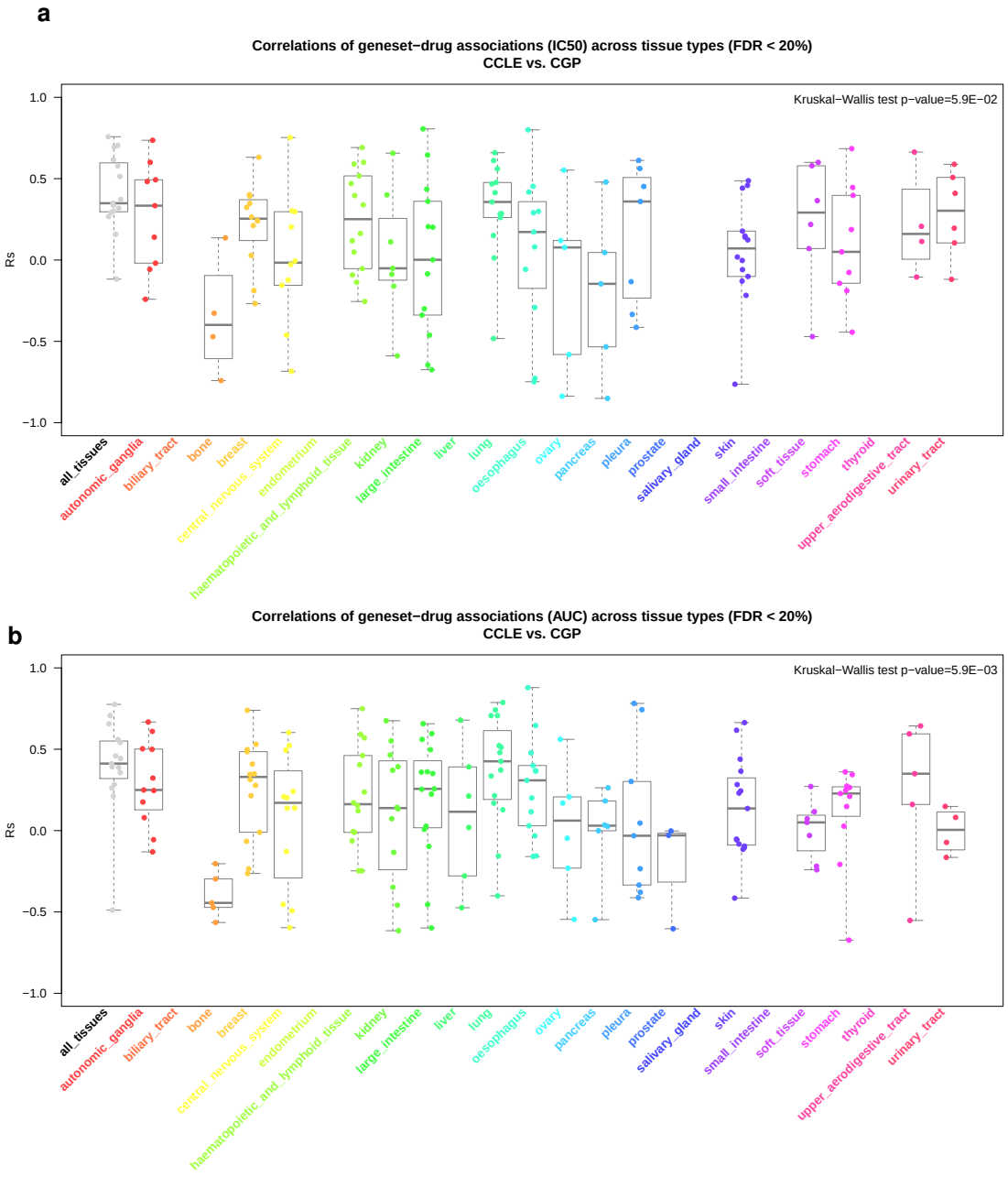
Supplementary Figure 15 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



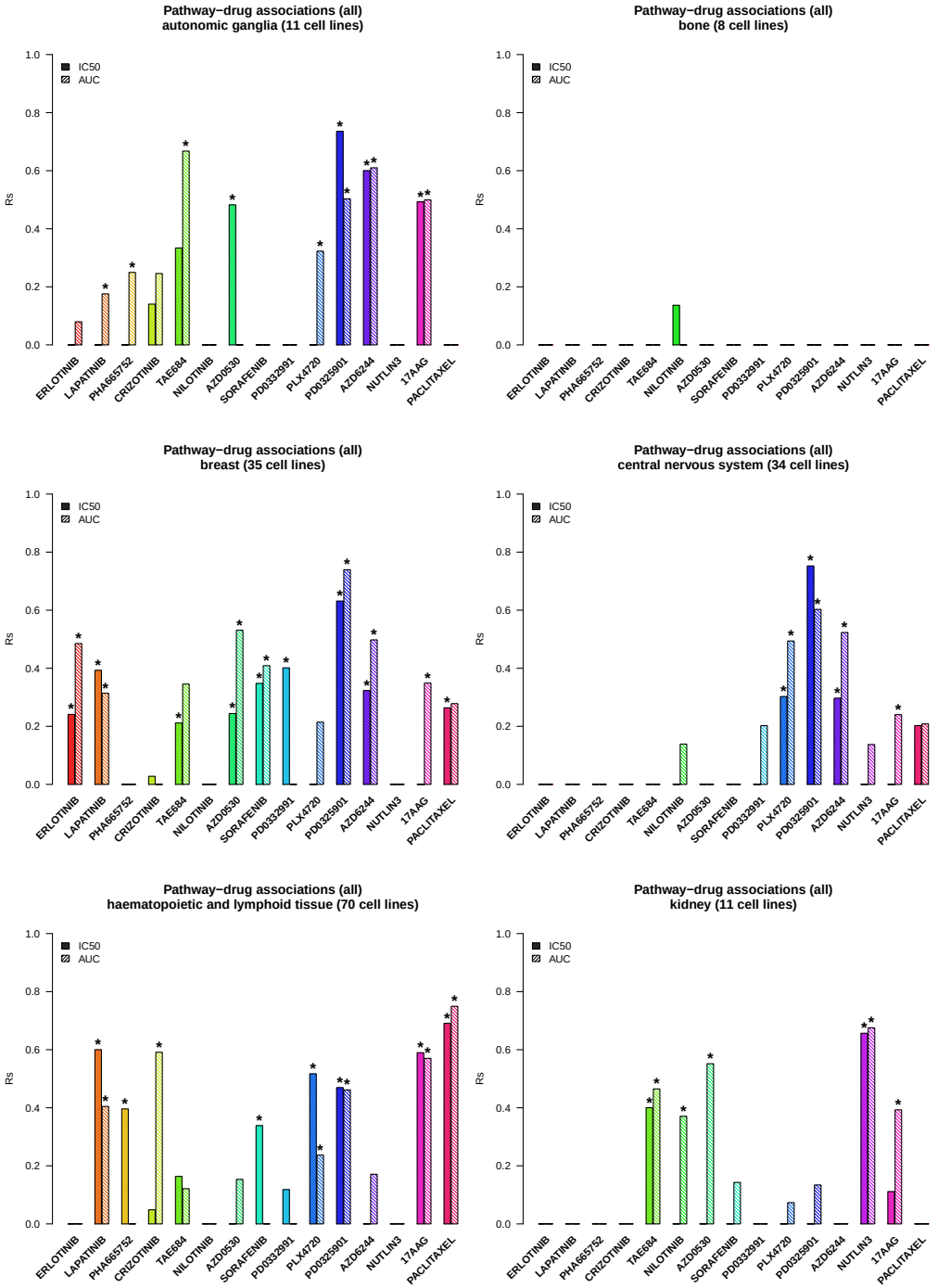
Supplementary Figure 16 (a) Scatter plots reporting the significant (FDR<20%) pathway-drug associations computed with IC_{50}, as quantified by the enrichment score from gene set enrichment analysis, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE. The last bar plot (bottom right corner) reports the positive Spearman correlation coefficient (R_s) for each drug.



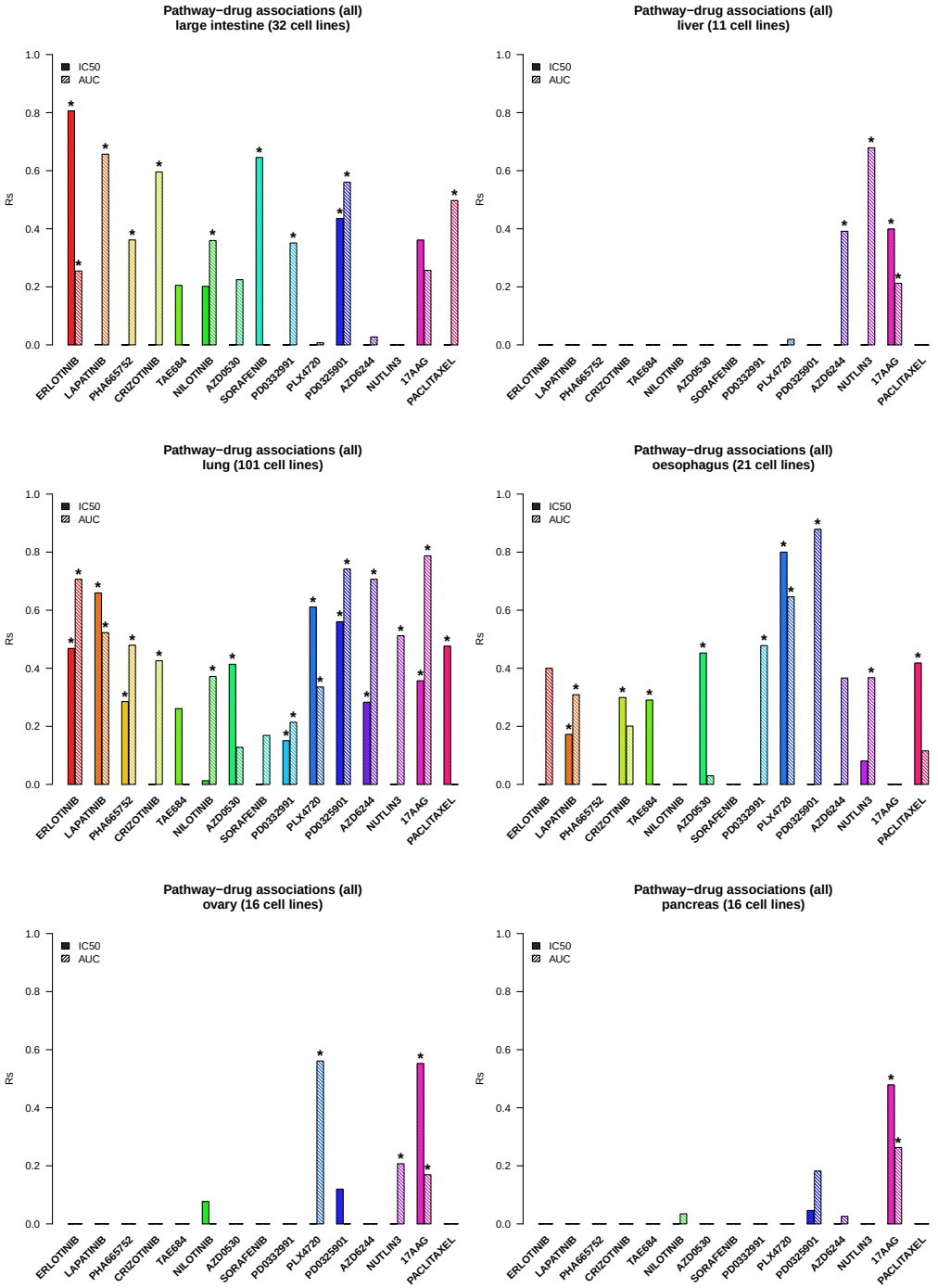
Supplementary Figure 17 (a) Scatter plots reporting the significant (FDR<20%) pathway-drug associations computed with AUC, as quantified by the enrichment score from gene set enrichment analysis, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE. The last bar plot (bottom right corner) reports the positive Spearman correlation coefficient (R_s) or each drug.



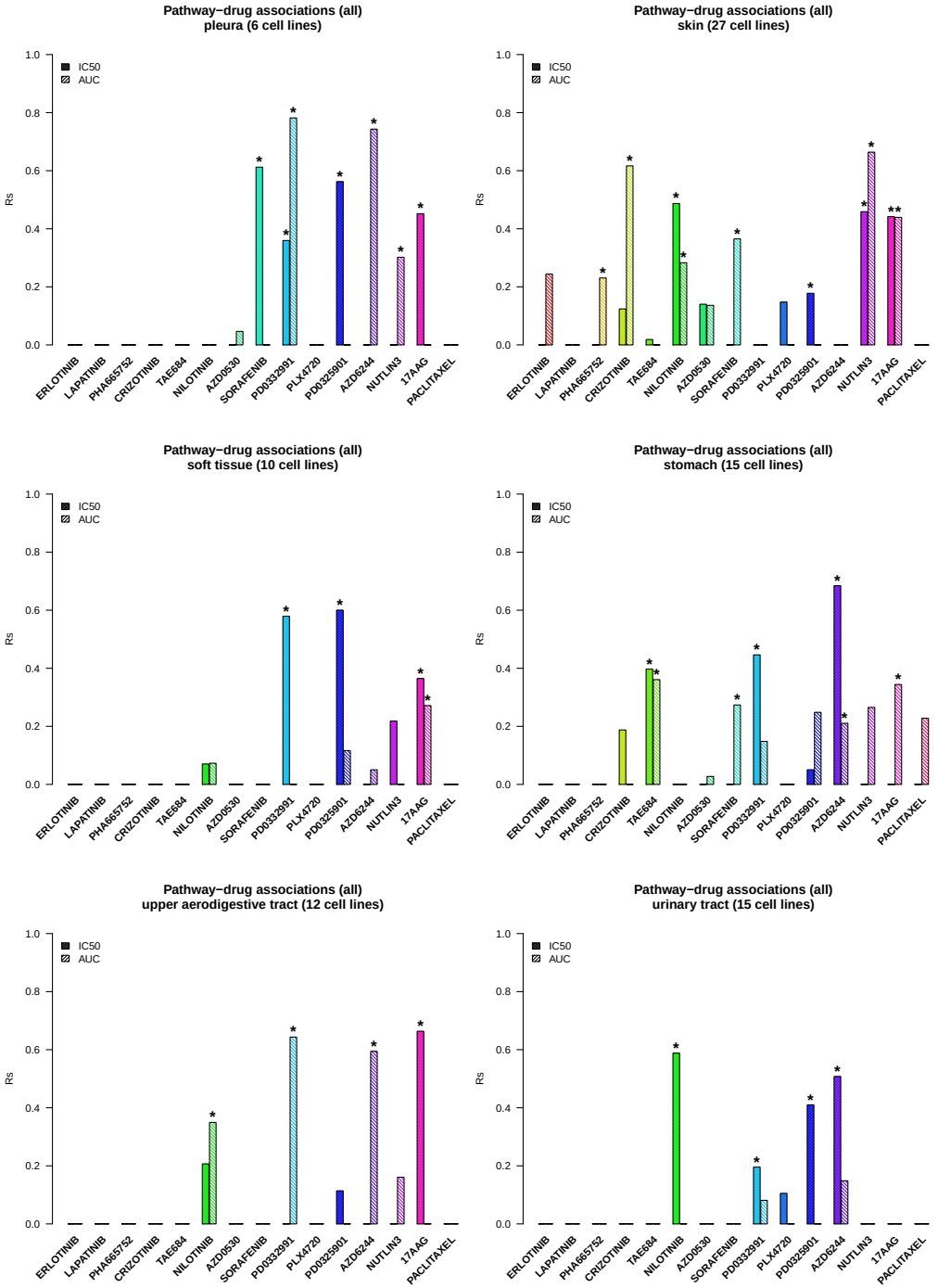
Supplementary Figure 18 Box plot of the correlations of the significant (FDR $<$ 20%) pathway-drug associations for the 15 drugs, across tissue types. (a) Correlations between significant pathway-drug associations computed with IC_{50} in CGP and CCLE; (b) correlations between significant pathway-drug associations computed with AUC in CGP and CCLE. Correlations were estimated using the Spearman coefficient (R_s). Kruskal-Wallis test was used to test whether correlations significantly depended on tissue type (upper right corner).



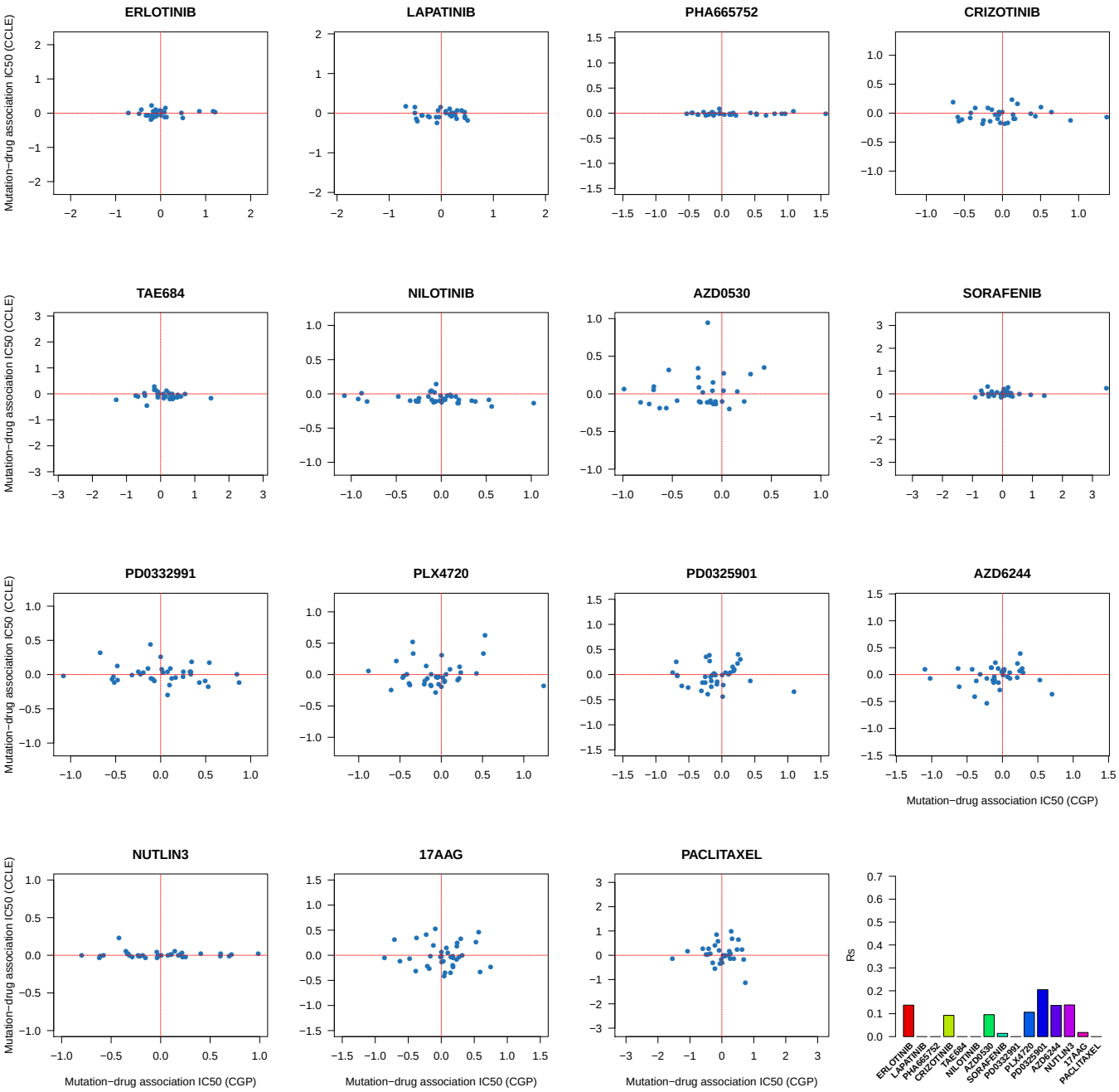
Supplementary Figure 19 Bar plot reporting the positive Spearman correlation coefficients (R_s) for significant (FDR $<$ 20%) pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value $<$ 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



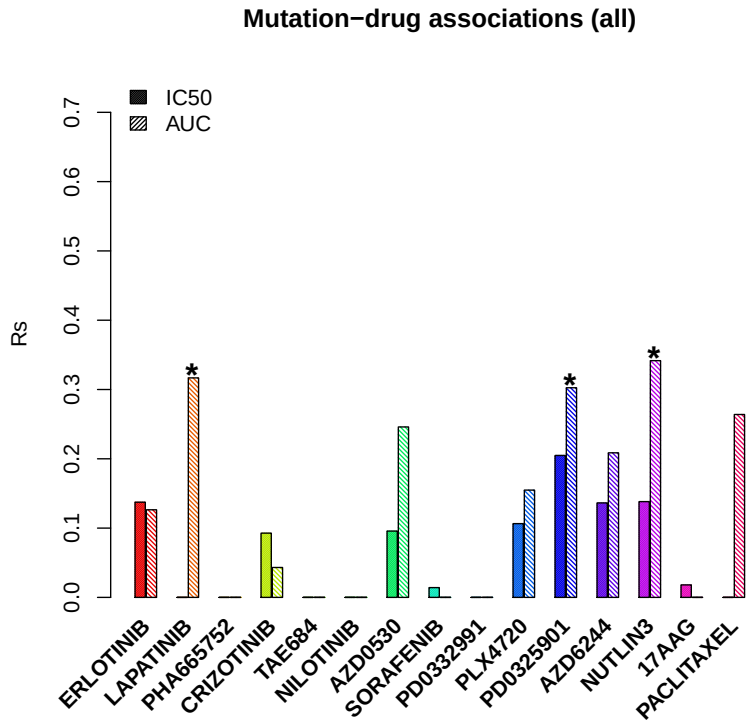
Supplementary Figure 19 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for significant (FDR $< 20\%$) pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



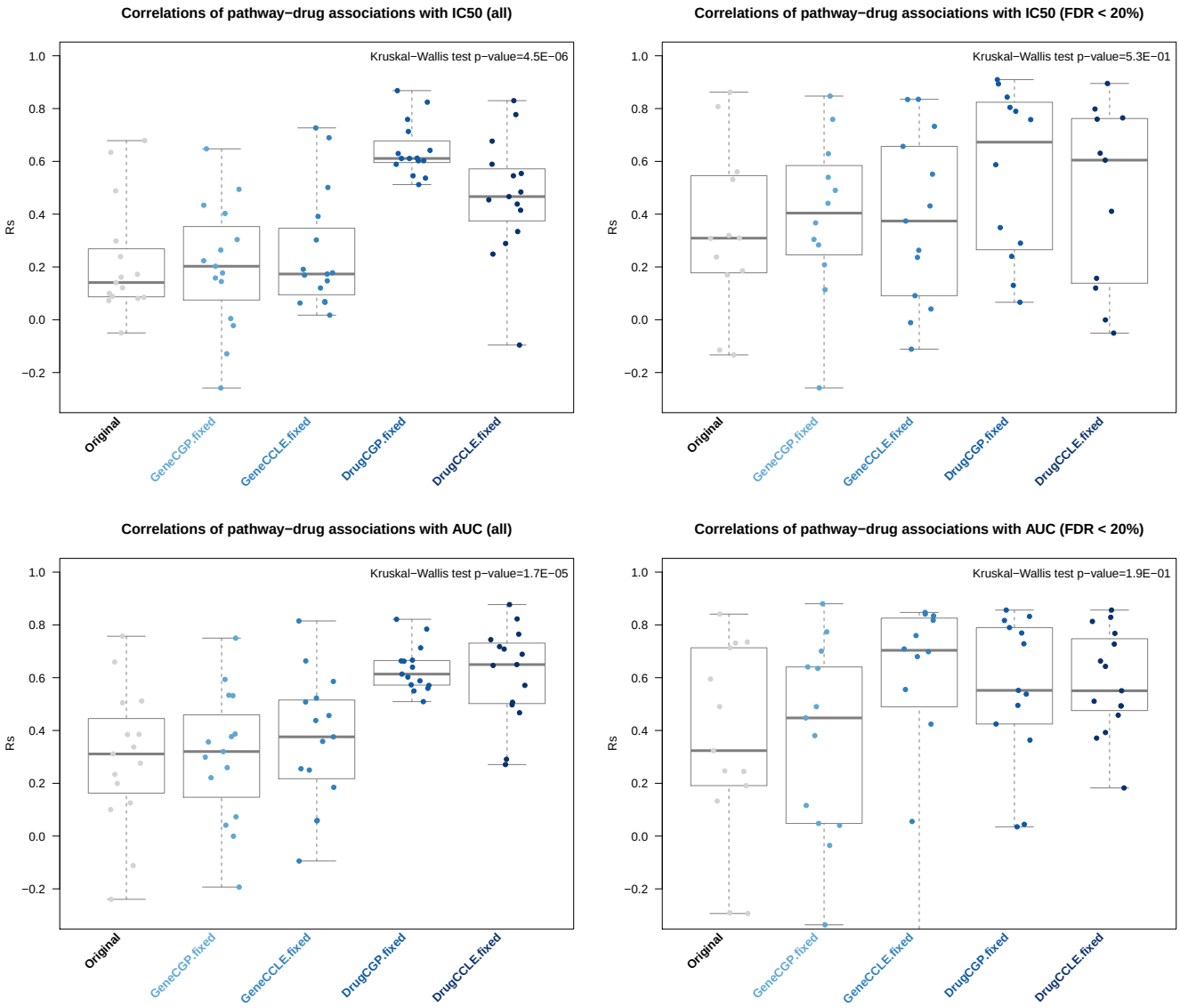
Supplementary Figure 19 (cont'd) Bar plot reporting the positive Spearman correlation coefficients (R_s) for significant (FDR $< 20\%$) pathway-drug associations computed with IC_{50} and AUC in CGP and CCLE, across tissue types. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05. If none positive correlations can be computed for a given tissue type, the plot is omitted.



Supplementary Figure 20 Scatter plots reporting the mutation-drug associations computed with IC_{50}, as quantified by the standardized coefficient of the gene of interest in a linear model controlled for tissue type, in the 471 cell lines and for each the 15 drugs investigated both in CGP and CCLE. The last bar plot (bottom right corner) reports the positive Spearman correlation coefficient (R_s) for each drug.



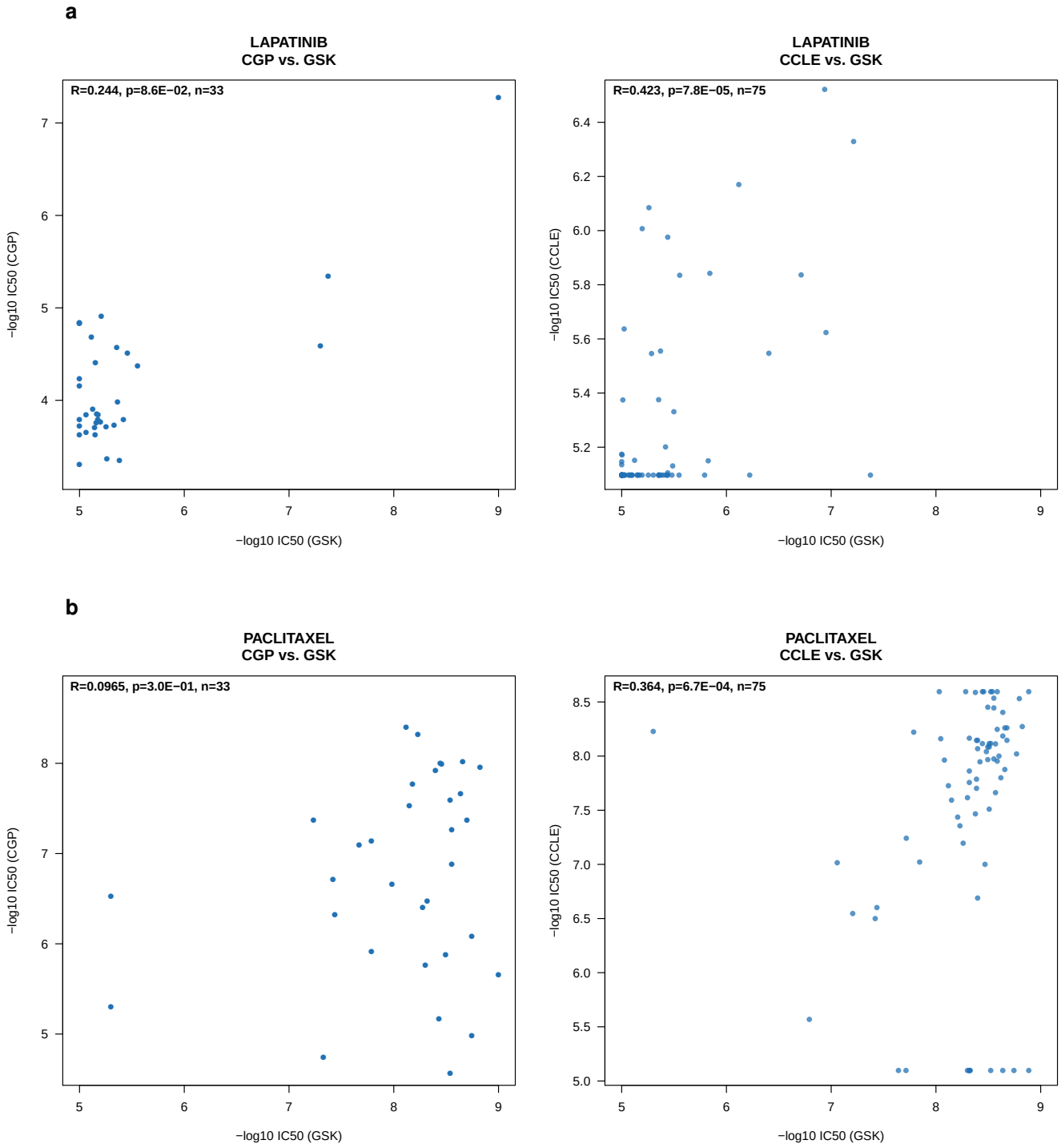
Supplementary Figure 21 Bar plot reporting the positive Spearman correlation coefficients (R_s) for the mutation-drug associations computed with IC_{50} and AUC measures both in CGP and CCLE. Significance of each correlation coefficient is reported using the symbol '*' if p-value < 0.05.



Supplementary Figure 22 Box plots reporting, for the 15 drugs in the 471 cell lines investigated both in CGP and CCLE, the correlations between the pathway-drug associations with IC_{50} and AUC, as well as the significant (FDR $<$ 20%) pathway-drug associations with IC_{50} and AUC. Each box represent the datasets used to compute correlations:

- 'Original' refers to the original datasets which are [CGP_g + CGP_d] vs. [$CCLE_g$ + $CCLE_d$],
- 'GeneCGP.fixed' refers to [CGP_g + CGP_d] vs. [CGP_g + $CCLE_d$],
- 'GeneCCLE.fixed' refers to [$CCLE_g$ + CGP_d] vs. [$CCLE_g$ + $CCLE_d$],
- 'DrugCGP.fixed' refers to [CGP_g + CGP_d] vs. [$CCLE_g$ + CGP_d],
- 'DrugCCLE.fixed' refers to [CGP_g + $CCLE_d$] vs. [$CCLE_g$ + $CCLE_d$].

where $_g$ and $_d$ stand for gene expressions and drug sensitivities, respectively. Kruskal-Wallis test was used to test whether correlations significantly depended on dataset (upper right corner).



Supplementary Figure 23 Scatter plots reporting the drug sensitivity measurements (IC_{50}) of (a) Lapatinib and (b) Paclitaxel in the 194 cancer cell lines screened in CGP, CCLE and GSK datasets.

References

[22] P E Andreotti, I A Cree, C M Kurbacher, D M Hartmann, D Linder, G Harel, I Gleiberman, P A Caruso, S H Ricks, and M Untch. Chemosensitivity testing of human tumors using a microplate adenosine triphosphate luminescence assay: clinical correlation for cisplatin resistance of ovarian carcinoma. *Cancer Res*, 55(22):5276–5282, November 1995.

[23] W T Bellamy. Prediction of response to drug therapy of cancer. A review of in vitro assays. *Drugs*, 44(5):690–708, November 1992.

[24] Joslyn K Brunelle and Baolin Zhang. Apoptosis assays for quantifying the bioactivity of anticancer drug products. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy*, 13(6):172–179, December 2010.

[25] Harold J. Burstein, Pamela B Mangu, Mark R Somerfield, Deborah Schrag, David Samson, Lawrence Holt, Debra Zelman, Jaffer A Ajani, and American Society of Clinical Oncology. American Society of Clinical Oncology clinical practice guideline update on the use of chemotherapy sensitivity and resistance assays., August 2011.

[26] Grace Ka Yan Chan, Tracy L Kleinheinz, David Peterson, and John G Moffat. A simple high-content cell cycle assay reveals frequent discrepancies between cell number and ATP and MTS proliferation assays. *PloS one*, 8(5):e63583, 2013.

[27] Ian A Cree. Chemosensitivity and chemoresistance testing in ovarian cancer. *Current opinion in obstetrics & gynecology*, 21(1):39–43, February 2009.

[28] S P Crouch, R Kozlowski, K J Slater, and J Fletcher. The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *Journal of immunological methods*, 160(1): 81–88, March 1993.

[29] L Galluzzi, S A Aaronson, J Abrams, E S Alnemri, D W Andrews, E H Baehrecke, N G Bazan, M V Blagosklonny, K Blomgren, C Borner, D E Bredesen, C Brenner, M Castedo, J A Cidlowski, A Ciechanover, G M Cohen, V De Laurenzi, R De Maria, M Deshmukh, B D Dynlacht, W S El-Deiry, R A Flavell, S Fulda, C Garrido, P Golstein, M-L Gougeon, D R Green, H Gronemeyer, G Hajnóczky, J M Hardwick, M O Hengartner, H Ichijo, M Jäättelä, O Kepp, A Kimchi, D J Klionsky, R A Knight, S Kornbluth, S Kumar, B Levine, S A Lipton, E Lugli, F Madeo, W Malomi, J-C W Marine, S J Martin, J P Medema, P Mehlen, G Melino, U M Moll, E Morselli, S Nagata, D W Nicholson, P Nicotera, G Nuñez, M Oren, J Penninger, S Pervaiz, M E Peter, M Piacentini, J H M Prehn, H Puthalakath, G A Rabinovich, R Rizzuto, C M P Rodrigues, D C Rubinsztein, T Rudel, L Scorrano, H-U Simon, H Steller, J Tschopp, Y Tsujimoto, P Vandenabeele, I Vitale, K H Vousden, R J Youle, J Yuan, B Zhivotovsky, and G Kroemer. Guidelines for the use and interpretation of assays for monitoring cell death in higher eukaryotes. *Cell death and differentiation*, 16(8):1093–1107, August 2009.

[30] Oliver Kepp, Lorenzo Galluzzi, Marta Lipinski, Junying Yuan, and Guido Kroemer. Cell death assays for drug discovery. *Nat Rev Drug Discov*, 10(3):221–237, March 2011.

[31] Silvia Miret, Els M De Groene, and Werner Klaffke. Comparison of in vitro assays of cellular toxicity in the human hepatic cell line HepG2. *Journal of biomolecular screening*, 11(2): 184–193, March 2006.

[32] Andrew L Niles, Richard A Moravec, and Terry L Riss. Update on in vitro cytotoxicity assays for drug development. *Expert opinion on drug discovery*, 3(6):655–669, June 2008.

[33] Andrew L Niles, Richard A Moravec, and Terry L Riss. In vitro viability and cytotoxicity testing and same-well multi-parametric combinations for high throughput screening. *Current chemical genomics*, 3:33–41, 2009.

[34] R D Petty, L A Sutherland, E M Hunter, and I A Cree. Comparison of MTT and ATP-based assays for the measurement of viable cell number. *Journal of bioluminescence and chemiluminescence*, 10(1):29–34, January 1995.

[35] Terry L Riss and Richard A Moravec. Use of multiple assay endpoints to investigate the effects of incubation time, dose of toxin, and plating density in cell-based cytotoxicity assays. *Assay and drug development technologies*, 2(1):51–62, February 2004.

[36] Terry L Riss, Richard A Moravec, and Andrew L Niles. Cytotoxicity testing: measuring viable cells, dead cells, and detecting mechanism of cell death. *Methods in molecular biology (Clifton, N.J.)*, 740:103–114, 2011.

[37] Venil N Sumantran. Cellular chemosensitivity assays: an overview. *Methods in molecular biology (Clifton, N.J.)*, 731:219–236, 2011.

[38] Nicola Tolliday. High-throughput assessment of Mammalian cell viability by determination of adenosine triphosphate levels. *Current protocols in chemical biology*, 2(3):153–161, September 2010.

[39] Engin Ulukaya, Ferda Ozdikicioglu, Arzu Yilmaztepe Oral, and Meral Demirci. The MTT assay yields a relatively lower result of growth inhibition than the ATP assay depending on the chemotherapeutic drugs tested. *Toxicology in vitro : an international journal published in association with BIBRA*, 22(1):232–239, February 2008.

[40] Jörg Weyermann, Dirk Lochmann, and Andreas Zimmer. A practical note on the use of cytotoxicity assays. *International journal of pharmaceutics*, 288(2):369–376, January 2005.

[41] D Wlodkowic, J Skommer, and Z Darzynkiewicz. Cytometry of apoptosis. Historical perspective and new advances. *Experimental oncology*, 34(3):255–262, October 2012.

[42] Donald Wlodkowic, Joanna Skommer, Chris Hillier, and Zbigniew Darzynkiewicz. Multiparameter detection of apoptosis using red-excitable SYTO probes. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 73(6):563–569, June 2008.